2022

Emergence Probabilities of COVID-19 Lineages and Motivations for Further Indigenous COVID-19 Research

Taylor Marcus

Emergence Probabilities of COVID-19 Lineages and Motivations for Further Indigenous COVID-19 Research

Abstract

At this point in the COVID-19 pandemic, there are vaccines available that can protect us against the virus. These vaccines work by creating antibodies that attach to spike proteins on the virus' surface, inhibiting them from binding the angiotensin-converting enzyme-2 (ACE2) receptor on our cells. To date, all COVID-19 lineages infect our cells by attaching to these ACE-2 receptors. This research project presents probability estimates of a new COVID-19 lineage emerging that binds a receptor other than ACE-2. This research also discusses available research on COVID-19 in Indigenous communities and why Indigenous research is so important. The methods used in this study involve a literature review, MATLAB coding, and self-reflection. It was found that a 95% chance of a new COVID-19 lineage emerging that binds a receptor other than ACE-2 over the next decade is consistent with the observed data to date. This observation would render our current COVID-19 vaccines ineffective and could revert us to a pre-vaccine stage in the pandemic. This would have major consequences for Indigenous communities, as they have suffered from COVID-19 more than the general population. Despite this, Indigenous Peoples have remained strong by implementing coping strategies based on their own culture and traditions. Further Indigenous COVID-19 research is needed to support the inequities present in Indigenous communities but also recognize their strengths.

Keywords

COVID-19 pandemic, Indigenous communities, COVID-19 lineages, Indigenous research

Motivation

My research is the result of my experiences and motivations as an Indigenous scholar. I grew up in the town of Chatham, Ontario as Métis and I never had a strong connection to my Indigenous culture. It wasn't until recently that I've felt an urge to find my Indigenous identity. At this point in my life, I've had four years of experience conducting scientific research from a euro-western framework, and I've become more motivated to learn about Indigenous research paradigms. The Head & Heart program was the perfect opportunity for me to do this.

I decided on my specific research project with the hopes of incorporating my supervisor's expertise, my own interests, and Indigenous perspective. My research background is in the department of medical biophysics, with emphasis on diagnostic imaging. I was fortunate enough to get paired with an incredible mentor, Dr. Lindi Wahl, whose background is in applied math. Her lab focusses on modelling and predicting the evolution of bacteria and viruses. Together, we decided that my research would focus on modelling different emergence probabilities of new COVID-19 lineages and discuss how COVID-19 may have affected Indigenous communities.

I am invested in this research because I absolutely love learning new skills. I was able to use my coding knowledge to learn how to model probability generating functions and predict future events. This research is also important to me because it is relevant to everyone worldwide. The

pandemic is something that has affected all people, but in different ways. Since the pandemic has affected me so significantly these past two years, I wondered how it may have affected some Indigenous communities. I am passionate about opening the minds of others, promoting empathy, and encouraging people to consider the challenges of marginalized groups. It is my hope that I can apply the knowledge and skills I have learned throughout my research to my future career as a physician, to promote decolonization and indigenization, and embody the values of equity, diversity, and inclusion.

Learning a new skill

To begin my learning journey and start my research, I had to learn about probability generating functions and how to create them in MATLAB. Dr. Wahl set me up with many great resources and past lectures of hers to learn more about these topics. She also presented me with a research paper that was previously conducted in her lab titled *Estimating the risk of pandemic avian influenza*.[1] This research paper used probability generating functions to estimate the probabilities of a pandemic emerging by the transfer of the influenza virus from birds to humans. Ironically, this study was conducted just before the start of the COVID-19 pandemic. To learn more about modelling probability generating functions, I recreated all the figures in this research paper before applying my new skill to my new research idea.

Applying the skill to a new idea

After recreating the figures from the previous paper, I felt confident to try and apply this knowledge to a new scenario. Since the COVID-19 pandemic is currently relevant worldwide, I wanted to estimate emergence probabilities of new COVID-19 lineages.

Viruses such as COVID-19 mutate over time to create new lineages. By definition, a lineage is a group of closely genetically related variants that emerged from a common ancestor. A variant is the same coronavirus but differs by one or more mutations, some of which we know are omicron and delta. Since these variants have different mutations, this affects how the virus behaves and infects humans.

I thought it would be interesting to graph the number of COVID-19 lineages against number of cases to evaluate their relationship. I found this information by first visiting Our World in Data website to determine the number of cases each month since the start of the pandemic in December 2019.[2] I then searched through cov-lineages.org to count the number of lineages that emerged in each of these months.[3] Using this data, I created a plot in MATLAB (Mathworks, R2020a, Natick, MA) of the number of lineages versus number of cases (Figure 1).

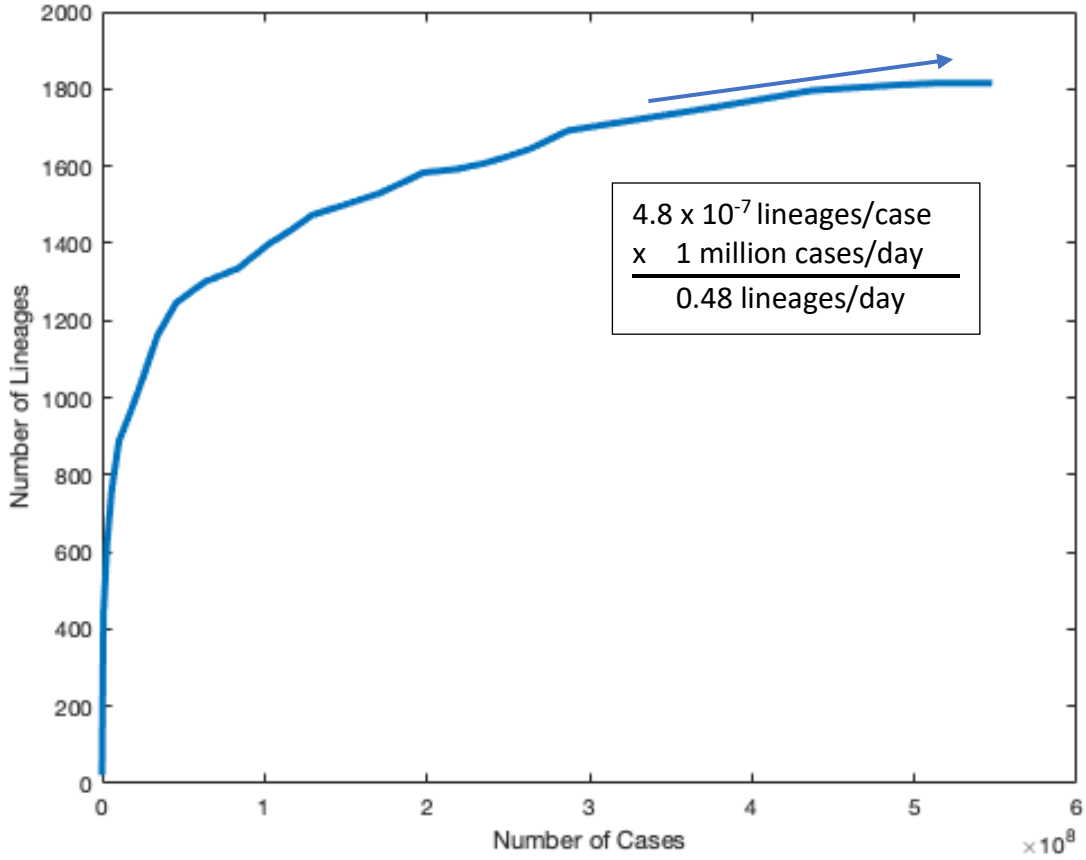

Figure 1: Number of COVID-19 lineages emerging with number of cases. The blue arrow represents the current slope or rate of emergence of new lineages per case. The text box shows the calculation of the number of lineages emerging per day based on the current rate of lineages emerging per case and the current number of cases per day.

Based on this plot, it appears that the number of lineages was increasing at an intense rate as cases grew from zero to 100 million, but recently the number of lineages emerging has begun to slow down. Based on the calculations in Figure 1, roughly one new lineage is emerging every two days. However, it is important to note that this data is only based on those individuals that went to get tested during their infection.

Thankfully, at this point in the pandemic, we have vaccines to protect us against viral infection. During COVID-19 infection, the spike proteins on the coronavirus attach to the angiotensin converting enzyme 2 (ACE-2) receptors on our cells (Figure 2a). In a vaccinated person, the vaccine induced the production of antibodies that can attach to these proteins and inhibit binding and therefore infection (Figure 2b). To date, all COVID-19 lineages enter our cells by attaching to these ACE-2 receptors.[4] But what would happen if a new COVID-19 lineage emerged that targeted a receptor other than ACE-2? Our antibodies would not be able to inhibit binding and therefore our vaccines would no longer be protecting us against these new COVID-19 variants.

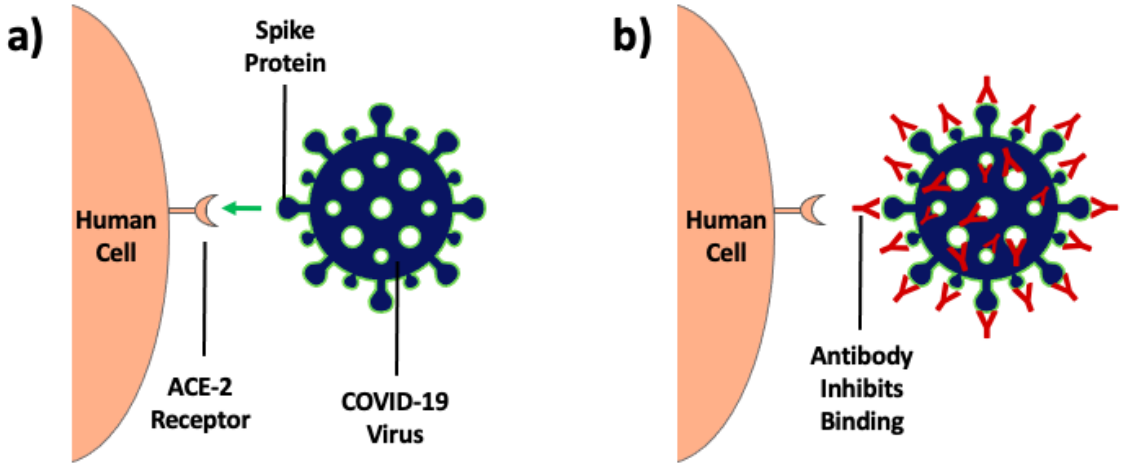

Figure 2: The COVID-19 virus attempting to infect a human lung cell a) without and b) with exposure to a vaccine. Normally, the spike proteins on the virus will bind to the ACE-2 receptors on the cell's surface. After exposure to a vaccine, the body produces antibodies that can bind to the spike proteins on the virus' surface and inhibit binding and therefore infection.

To date, there have been a total of 1815 distinct COVID-19 lineages. I let p_0 represent the probability that one individual lineage would evolve to target a receptor other than ACE-2. By making the conservative assumption that the fate of each lineage is independent, I will be estimating an upper bound of the emergence of a lineage that evolves to bind a new receptor. For each individual lineage, the lineage continues to bind the ACE-2 receptor with probability $1-p_0$ or evolves to bind a new receptor with probability p_0. If N COVID-19 lineages have occurred worldwide, and each of them continue to bind the ACE-2 receptor with probability $1-p_0$, the probability of observing N lineages that all continue binding ACE-2 is given by the following expression:

$$Q(p_0, N) = (1 - p_0)^N$$

Figure 3 plots Q (p_0, N) as a function of p_0 for various N values. The bold purple line shows results for N=1815 (the number of COVID-19 lineages in humans as of July 2022). The horizontal line (Q^*) represents a 5% rejection threshold, where I assume that the parameters are statistically insignificant with observed data if the observation is less than 5%. For N=1815, I see that the model is consistent with observation if $p_0 \leq 0.00165$. This value is shown as a filled in circle in Figure 3 and represents the upper bound on the probability that an individual lineage will evolve to bind a new receptor. This upper bound will now be referred to as $\bar{p}_0$.

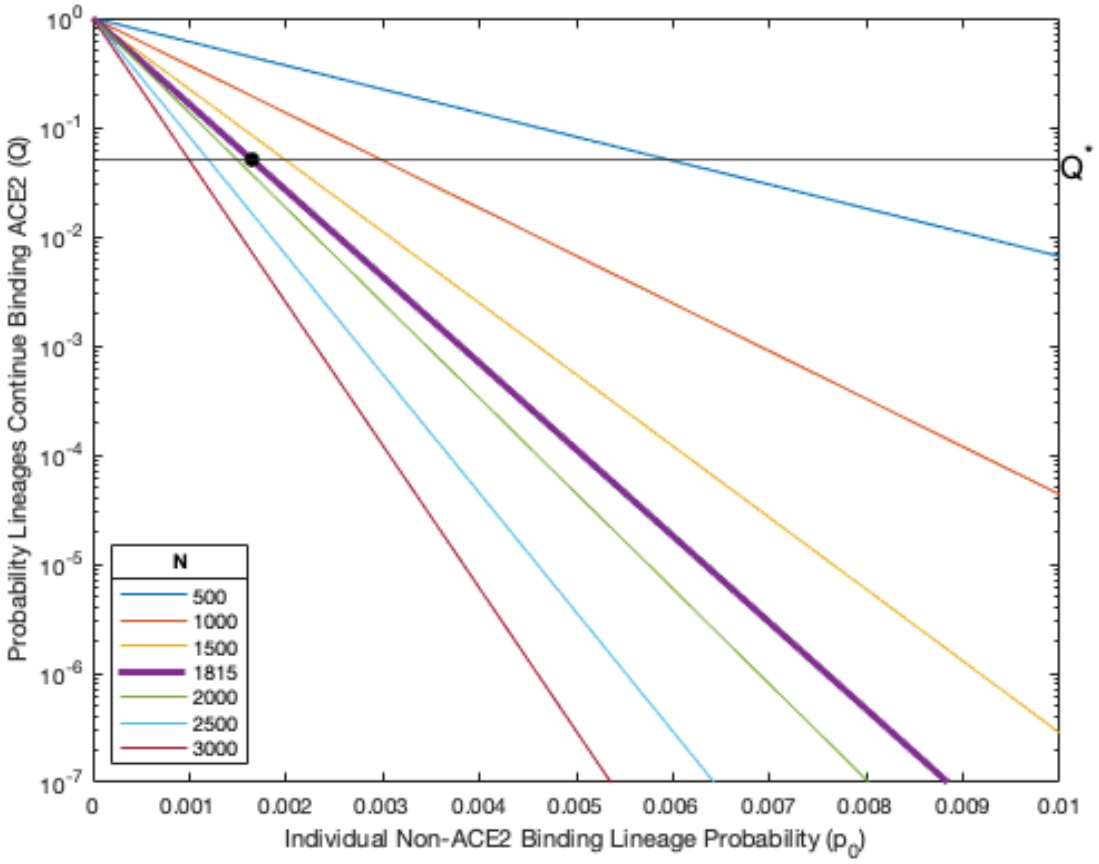

Figure 3: The probability of observing N lineages that continue to bind the ACE-2 receptor, Q, as a function of the emergence probability of an individual lineage binding a new receptor, p_o. Each line plots $Q(p_0,N)$ for a different value of N. The horizontal line shows the rejection threshold (Q^*=0.05), while the bold line shows Q for N_c=1815, the current worldwide total number of lineages. The upper bound p_0 which is consistent with observed data is shown by the filled circle ($\bar{p}_0$=0.00165).

The figure demonstrates that if the probability of lineages continuing to bind to ACE-2 is high, the individual probability that a lineage emerges to bind a new receptor is extremely low. As the number of lineages that continue binding ACE-2 increases, the statistical estimate of the upper bound on p_0 will be reduced. If we have observed 1815 COVID-19 lineages, none of which have bound a new receptor, for rejection threshold Q^*, the upper bound on p_0 that is consistent with data is given by:

$$\bar{p}_0 = 1 - (Q^*)^{\frac{1}{N}}$$

Figure 4 illustrates how this estimate of $\bar{p}_0$ would change as more new lineages are observed in the future, assuming they all continue binding the ACE-2 receptor. Using the 5% rejection threshold and the large number of lineages that all bind ACE-2, $\bar{p}_0$ is insensitive to N.

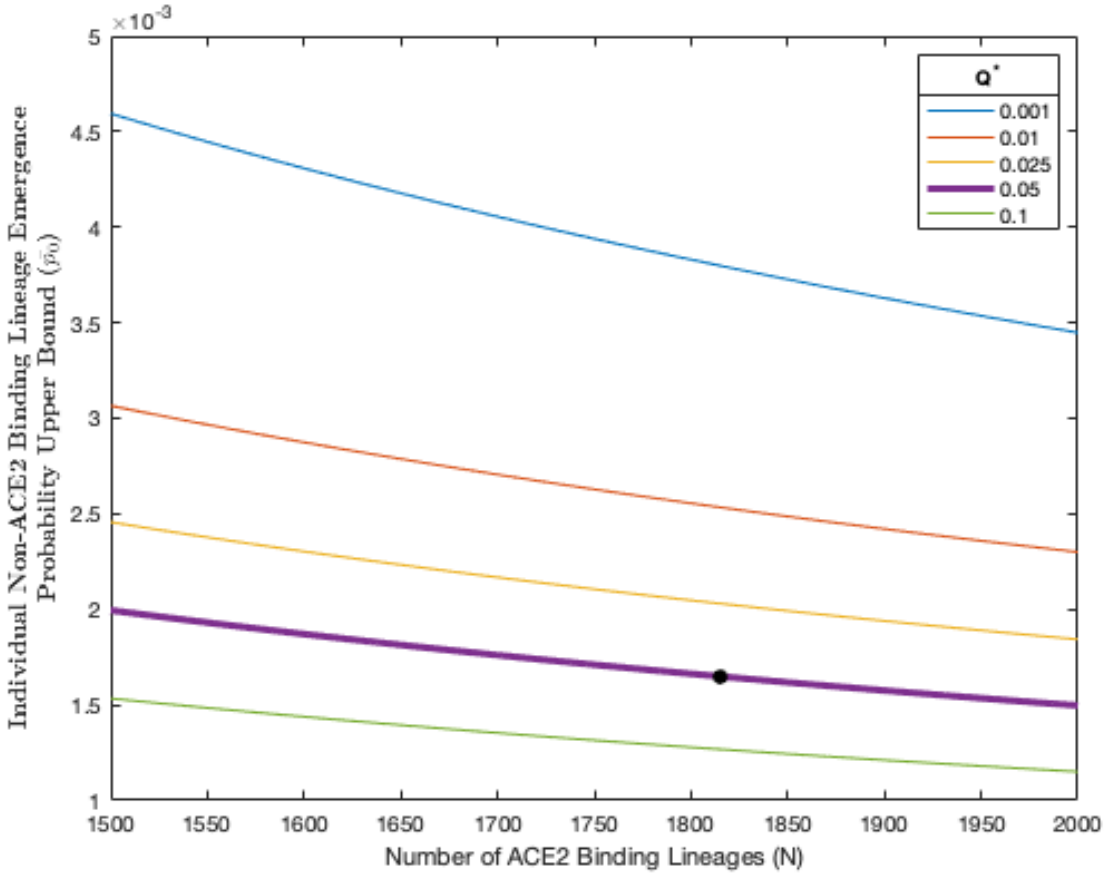

Figure 4: The upper bound estimate of the probability an individual lineage evolves to bind a receptor other than ACE-2 as a function of number of observed lineages. Each line plots $\bar{p}_0$ for a different value of the rejection threshold Q^*. The bold line shows $\bar{p}_0$ for the 5% rejection threshold. The upper bound on p_0 which is consistent with currently observed data is shown by the filled circle ($\bar{p}_0$=0.00165, N=1815).

Using the rate of emergence of new COVID-19 lineages per day as found by using Figure 1, we notice that the rate of emergence of new lineages per year is roughly 174.7. With the simplest assumption of a time-homogenous Poisson process, we can expect an average of *rn* further lineages to occur in *n* years. If we've observed N_c=1816 lineages to date, we can expect that our estimate of $\bar{p}_0$ could be reduced at best to the following equation, in n years, if we assume all lineages continue binding the ACE-2 receptor in that interval.

$$\bar{p}_0(t) = 1 - (Q^*)^{\frac{1}{N_c+rn}}$$

Figure 5 illustrates how the upper bound estimate of the probability of an individual lineage evolving to bind a new receptor will change over the next *n* years, if we continue to observe new

lineages emerging at the same rate. Based on the plot, the numerical value of $\bar{p}_0$ is expected to decrease but remains relatively stable over time.

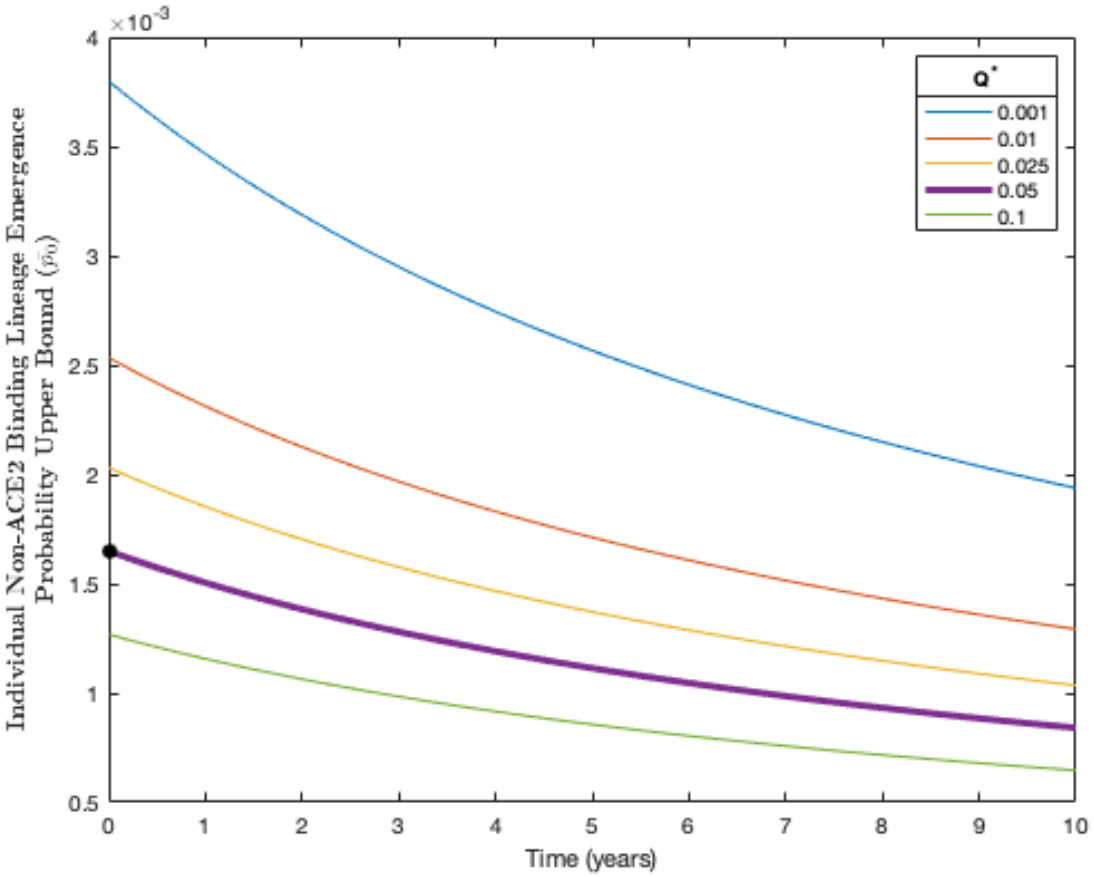

Figure 5: The upper bound estimate of the probability an individual lineage evolves to bind a receptor other than ACE-2 as a function of time, assuming the rate of lineages emerging per year remains the same at r=174.7 new lineages per year. Each line plots $\bar{p}_0$ for a different value of the rejection threshold Q^*. The bold line shows $\bar{p}_0$ for the 5% rejection threshold. The upper bound on p_0 which is consistent with currently observed data is shown by the filled circle ($\bar{p}_0$=0.00165, time=0).

Using this value of $\bar{p}_0$, I can predict the cumulative probability that a new COVID-19 lineage will emerge that binds a receptor other than ACE-2 over the next decade. Once again, assuming a time-homogenous Poisson process, it is expected that *rn* new lineages emerge on average in the next *n* years. If $\bar{p}_0$ is the upper bound, a lower bound on the probability that all lineages continue to bind the ACE-2 receptor over the next n years would be $Q(\bar{p}_0,N)=(1-\bar{p}_0)^{rn}$. Overall, an upper bound on the probability that one or more lineages evolve to target a new receptor would be given by:

$$E(n) = 1 - Q(\bar{p}_0, rn) = 1 - (1 - \bar{p}_0)^{rn}$$

Figure 6 plots the non-ACE-2 binding lineage emergence probability, E(n), as a function of n in years for various values of $\bar{p}_0$. The bold yellow line shows results for the upper bound previously derived, $\bar{p}_0$=0.00165. These results show that although the individual emergence probability for a lineage evolving to bind a new receptor is low, it is still possible that the cumulative probability

of a lineage emerging that binds a new receptor could be increasingly high. Alternatively, a 95% chance of a new COVID-19 lineage emerging that binds a receptor other than ACE-2 over the next decade is technically still consistent with the observed data to date. As this would render our current COVID-19 vaccines ineffective, it is important to be aware of this possibility.

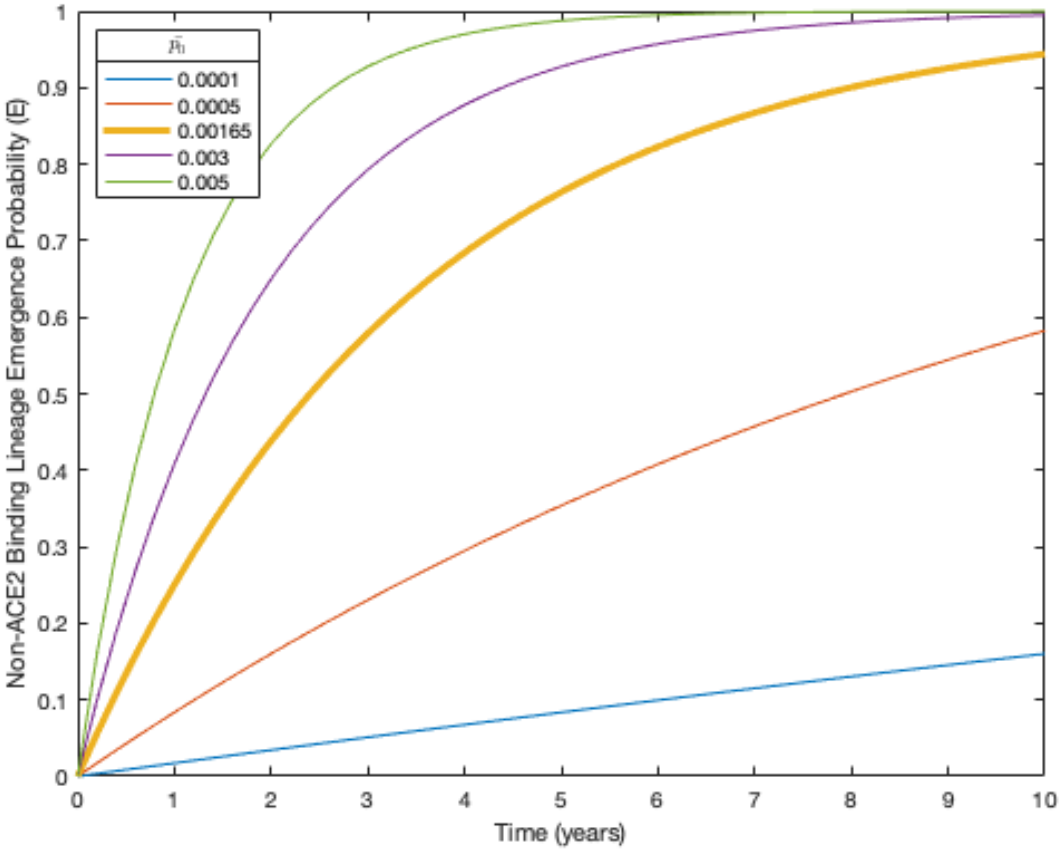

Figure 6: The probability of a non-ACE2 binding lineage emerging, E, versus time in years. Each line plots E(n) for a difference value of $\bar{p}_0$. The bold yellow line shows E for $\bar{p}_0$=0.00165, the estimate of this upper bound using current data.

Reflection

The probability estimates I've mentioned in this paper are incredibly interesting mathematical estimates, but we must keep in mind that we still can never be sure that an event will or will not happen in the future. What we do know for certain, is that the COVID-19 pandemic has affected all people worldwide, and in more ways than one. Businesses have shut down, schools have turned to remote learning, and gatherings have been limited. Not only has COVID-19 affected the physical health of many, but also their mental health, finances, and education, among many other things.

Indigenous communities specifically have suffered throughout this pandemic. Indigenous Peoples on average live shorter lives, have higher rates of disease and injury, and have higher

rates of hospitalizations for preventable conditions.[5] This is in part due to the healthcare inequities they experience, which stemmed from colonization. Since many Indigenous Peoples live on reserves that are rural or remote, this further contributes to their lack of access to quality healthcare. With this information, it's not surprising that during the current pandemic Indigenous communities have suffered more than the general population. COVID-19 has disrupted Indigenous cultural practices and social distancing has hindered our expression of relationality. These points are paramount, as cultural practices are our medicine.

It is important to point out the relevant hardships and disparities present in Indigenous communities, so that these inequities can be supported with adequate resources and services. However, I've learned a lot about Indigenous research this summer, and the ways that euro-western research often focusses on the negatives and what needs to be fixed. In the past, research has been done *on* Indigenous peoples, instead of *with* them. This research often included biases of Indigenous culture and positioned euro-western norms as superior. This is one reason why there is a need for Indigenous research - research that properly represents Indigenous perspectives and experiences. Indigenous research is not about extracting information and discovering something new, but it is a ceremony that improves your relationship with an idea.[6]

It is important to recognize that Indigenous communities are extremely resourceful and resilient, and during the difficulties of the pandemic, they have implemented their own strategies to successfully cope that are based on our culture, ways of being, knowing, and doing. Indigenous Peoples have further embraced their cultural practices to support their physical, mental, and emotional wellbeing. This includes activities that honour the land and the cosmos and remembering the importance of community. They have also embraced their traditional healing methods, which include sage, sweet grass, and other land-based medicines.[7] Although Indigenous Peoples have suffered deeply during this global crisis, it is important to acknowledge that there have been many ways that they have successfully responded to the COVID-19 pandemic.

To ensure Indigenous COVID-19 research is more relevant to communities in the future, I would love to visit Indigenous communities in Ontario, discuss the ways that the pandemic has negatively affected them, but also how they overcame these challenges as a community. During Indigenous research, it is important to immerse yourself in the community and build relationships. It is not enough to merely interview others but to share stories and perspectives in groups so that ideas can be elaborated and reflected on. Relationships are important not only between interviewer and participant, but between participants. For research to have the greatest impact, it must be easily accessible to and shared by others. This way, during the next pandemic, we will all know our strengths and weaknesses.

References

1. Tripathi A, Dhakal HC, Adhikari K, Chandra Timsina R, Wahl LM. Estimating the risk of pandemic avian influenza. *J Biol Dyn*. 2021;15(1):327-341. doi:10.1080/17513758.2021.1942570
2. Daily new confirmed COVID-19 cases. Our World in Data. https://ourworldindata.org/covid-cases. Accessed August 5, 2022.

3. Lineage List. cov-lineages.org. https://cov-lineages.org/lineage_list.html. Accessed August 5, 2022.
4. Ni W, Yang X, Yang D, et al. Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19. *Crit Care*. 2020;24(1):1-10. doi:10.1186/s13054-020-03120-0
5. Power T, Wilson D, Best O, et al. COVID-19 and Indigenous Peoples: An imperative for action. *J Clin Nurs*. 2020;29(15-16):2737-2741. doi:10.1111/jocn.15320
6. Wilson S. *Research Is Ceremony: Indigenous Research Methods*. Black Point, Nova Scotia: Fernwood Publishing; 2008.
7. Godin G. Impacts of COVID-19 on Indigenous Communities in Canada. *Heal Fit J Canada*. 2011;4(1):18-22.