When Corona Meets Darwin

When Corona Meets Darwin:

Exploring the Science Behind COVID-19

By
Zhaobin Xu and Dongqing Wei

Cambridge
Scholars
Publishing

When Corona Meets Darwin: Exploring the Science Behind COVID-19

By Zhaobin Xu and Dongqing Wei

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-0364-0531-1
ISBN (13): 978-1-0364-0531-1

With this book, I dedicate to those souls who have suffered from the pain of COVID-19, and to those science devotees who dare to pursue the truth.

—The Authors

CONTENTS

PREFACE

The COVID-19 pandemic has brought immense harm to human society, claiming many lives and unfolding as a tremendous tragedy. It has also deeply impacted global order and economic exchanges. From 2020 to the end of 2022, the exchange of personnel between China and the rest of the world was virtually interrupted. Moreover, it has challenged our understanding of infectious diseases, rendering many classic theories seemingly inadequate when faced with the novel coronavirus.

I am writing this book with two objectives in mind. The most important objective is to offer the general public a fresh perspective on COVID-19, one that differs from the reporting found in scientific journals. By reading this book, you will gain updated and profound insights into the interactions between the human immune system and viruses. This will help you understand the various perplexing phenomena observed during the COVID-19 pandemic, phenomena that experts have struggled to explain. The second objective of writing this book is to break down the so-called academic barriers. When real problems arise, people often discover how bureaucratic and inefficient the academic community can be. For example, when the COVID-19 pandemic surged, no one could accurately predict its development trends, explain the significant differences in infection symptoms, or efficiently produce effective medications. In contrast, we witnessed how the pandemic brought a massive windfall for researchers, with a plethora of mediocre articles being published in authoritative journals. Meanwhile, numerous prestigious medical journals like “The Lancet” saw a substantial increase in their impact factor with the arrival of COVID-19.

The academic community has now become a wasteland, where even articles published in Nature and Science are, by my estimation, 90% excellent examples of mediocrity, while the remaining 10% consists of content unrelated to my field, which I cannot comprehend nor evaluate. As a result, outstanding academic viewpoints fail to receive public attention through conventional academic channels, while scientific research occupying prestigious journals often turns out to be mediocre at best. These mediocre works lack sufficient innovation in their methods and are essentially the result of extensive human and financial resources, with little to no theoretical innovation. If Gregor Mendel's theory of hybridization were submitted to today's prestigious journals, it would most likely be rejected because great theories are always ahead of their time, and their credibility is inevitably questioned. Nowadays, authoritative journals, just like funding applications, prefer to accept reliable yet mediocre works rather than embrace forward-thinking theories. I believe that the general public is the sole driving force behind scientific progress, not the academic community. Just as improving the performance of certain products requires feedback from consumers, the generation of a great theory necessitates this powerful market foundation provided by the public. This book is aimed at the majority of people who have received a basic higher education. The language used is both profound and easy to understand. In addition to providing a clear understanding of our theories, each chapter is accompanied by appendices containing relevant scientific papers, allowing interested professionals to delve deeper and provide further references and comments.

Lastly, I hope that you will enjoy this book because it truly stands out among its peers!

Zhaobin Xu
Dongqing Wei

CHAPTER 1

THE ROLE OF MATHEMATICAL MODELS

> "A science is only in a perfected state when it can be expressed mathematically."
> —Karl Marx

Mathematics has played an extremely critical role in scientific progress. However, for ordinary people, mathematics is nothing more than a tool for passing exams. Besides simple arithmetic, mathematics has not played an important role for most people. On the other hand, advanced mathematical research often easily drifts away from practical applications and moves towards extreme number theory and abstract theory research. This can lead to the increasingly noticeable phenomenon of mathematics being detached from the market. In recent years, many prestigious mathematics awards, such as the Fields Medal, are known only by name, with no concrete understanding of their practical applications. Science that is disconnected from the public's understanding does not thrive. The progress of human science is not just the result of a few pioneering individuals' breakthroughs but rather the advancement of collective human thought. If science does not have the recognition of the public, then even the theory of gravity is merely a ridiculous hypothesis.

There are many reasons why mathematics has become detached from public understanding. One reason is that contemporary science, including mathematics, relies too heavily on the research achievements of predecessors. This leads to the phenomenon of increasingly limited research content. To illustrate

this point, suppose you're a high school student, your mathematical research would be studying a dog. If you're a college student, your mathematical research would be studying the tail of a dog. And if you're a doctoral student, your mathematical research tends to be a hair on a dog's tail. This creates a bad phenomenon where there are fewer and fewer disruptive and groundbreaking research projects. If the initial road is wrong, no matter how much progress you make later, you will not be able to get out of this dead end. Of course, there are many other reasons for mathematics becoming detached from the public. I personally believe that the most important reason is due to our rigid mathematics education. Even in developed Western countries with advanced educational philosophies, mathematics education is still flawed. The most significant mistake is that in science education, educators often do not emphasize the derivation of mathematical equations but instead overemphasize the application of certain physical formulas. For mathematics, the most important education should derive from how the equations are derived, such as the three fundamental laws of kinematics. Education should not focus just on how these laws are applied for intellectual tests or exams. This has led to public resistance towards mathematical research and it is not difficult to understand why a generally healthy person would not engage in mathematical research.

Before delving into the details of our research work, it is essential to first introduce the role of mathematical models and how they influence our understanding. The classic content we learn often does not involve the process of problem deduction. Most of the time, a simple mathematical formula is introduced for ease of understanding, while the most exciting and essential problem deduction process is often overlooked.

Let's take a simple example to illustrate this point. We will discuss Newton's laws of motion, a topic covered in secondary school physics. Most of us are familiar with these simple laws, which state that an object will move at a

constant velocity in the absence of external forces, and that the object's velocity will change when acted upon by a force, with the degree of change proportional to the force and inversely proportional to the object's mass. These principles seem straightforward because of the repeated results we have seen, but the actual deduction process is complex and fascinating.

The reason for its fascination lies in the fact that we cannot directly measure velocity. We can only measure time, weight, and length. For example, we can observe a car accelerating downhill on a slope, measure the angle of the slope, the length of the slope, and the time it takes for the car to slide down. However, we cannot measure the car's velocity, which led to the invention of calculus. If we express this in a mathematical formula, it is written as

$$\frac{dv}{dt} = \frac{F}{m};$$

$$\frac{dS}{dt} = v;$$

This was only a mathematical hypothesis at the time, and to prove that it was indeed following this mathematical equation, we had to establish a relationship between distance and time, as these were the two variables that we could directly measure. Establishing this relationship was challenging at the time, and Newton used calculus to deduce the classic formula as below:

$$\mathrm{S} = v_0 t + 0.5\frac{F}{m}t^2$$

The $\frac{dS}{dt} = v$ part of the formula is widely accepted, and the only dubious part is $\frac{dv}{dt} = \frac{F}{m}$. Through this integration calculation, Newton successfully linked distance and time, and by measuring the relationship between distance and time, he proved the existence of this relationship.

Looking back, our education at the time was mostly about rote learning, with no emphasis on the actual process of discovery. The wonder of scientific discovery lies in its foresight and prescience. If there were accurate measuring instruments for velocity and force at the time, Newton would not have appeared so great. The derivation of gravitation was similar.

For a long time, life sciences have been an empirical experimental science. Although in recent years, it has made great progress thanks to the advances in instrument equipment based on physical principles, such as high-throughput sequencing and cryo-electron microscopy, this has largely hindered its further development. One of the major drawbacks is the difficulty in comprehensively analyzing certain phenomena quantitatively.

A significant problem associated with the development of interdisciplinary fields related to life sciences is that researchers often lack knowledge in multiple disciplines. For fields related to viruses and immunology, many mathematical models of host-virus interactions were proposed as early as the 1980s. However, most of these models were based on parameter estimation through data fitting and cannot make systematic predictions. The root cause of this problem is that most modelers have a pure mathematical background. A model that relies solely on increasing parameters or compartments is not reliable because with enough parameters, a good fit can always be achieved. This is just like what the famous scientist von Neumann once said: "With four parameters I can fit an elephant, and with five, I can make it wiggle its trunk."

However, in the current academic community, as long as you have experimental data, no one cares about the correctness of your methods. Many formulas are purely based on mathematical fitting and lack physical mechanisms behind them. Reviewers only care about the fitting degree of your model and whether you have used fixed processes for analysis, such as balance point analysis, bifurcation analysis, and parameter sensitivity

analysis. Therefore, although these works have somewhat applications, they lack good grafting ability and wide-ranging predictive ability. Turning a large amount of experimental data into concrete model case reports is a huge failure in the development of applied mathematics.

In the following chapters, we will provide specific analysis and criticism of these common metaphysical academic studies. In conclusion, although theoretical biology research has risen greatly, there is still a lack of simulation work that can accurately reflect certain complex life processes in biological research.

The purpose of our book is to quantitatively study the development and changes of COVID-19 and scientifically evaluate a series of social policies based on COVID-19 through systematic mathematical modeling. My articles may offend many people, including academic authorities and various government officers. However, the correctness of the theory is not determined by these people, but by the public. I hope that my theory can trigger widespread discussions in society.

Before embarking on this brainstorming session, I must first introduce mathematical models and how to solve them. I assume that many readers do not have knowledge of calculus, but this does not affect your understanding of the content of the book. Let's take a simple example. Suppose we know Newton's laws of motion, and we want to analyze the free fall of an object without any resistance. Our goal is to determine the position and velocity of the object at time t after it is released. Since we know that the acceleration is g, the velocity at time t is simply gt. This is easy to understand, but expressing it as an equation gives us $\frac{dv}{dt} = g$. Using a numerical method, which is understandable by computers, how do we calculate the numerical value of gt? We use an iterative method, for example, to calculate the velocity at the third time point. We know that the velocity at the third time

point is derived from the velocity at the second time point, with a time interval of one and an acceleration. So, we first calculate the velocity at the second time point by adding the acceleration multiplied by the time interval to the velocity at the second time point. This way, we can calculate the velocity at the third time point. The same principle applies when calculating the velocity at the second time point using the velocity at the first time point, and using the initial velocity and acceleration to calculate the velocity at the first time point. By continuously iterating, we can calculate the velocity at the N^{th} time. For simpler cases, such as the one mentioned earlier where the acceleration is constant, we can directly use the formula $v = gt$ to solve for the velocity at any given time. For a more complex situation, such as calculating the relationship between descent distance and time, it may become slightly more complicated. We know that $\Delta s = v\Delta t$, but in this case, v is changing with time, $v = gt$. However, the iterative method is still effective. For example, if we want to calculate the descent distance at the third time point, we need to add the descent distance from the second time point to the third one on the basis of the descent distance at the second time point. We can assume that the velocity during a short period, from the second to the third time point, is constant. In this case, we only need to use the following formula to calculate the displacement at the third time point:

$$S_3 = S_2 + \mathrm{v}\Delta\mathrm{t}$$

The velocity we need to obtain is the average velocity during this time period, which is the velocity at 2.5 time point and can be calculated using $v = gt$. This way, we can calculate the displacement at any moment through continuous iteration. Some people may think that the method I described is very clumsy, why not simply use $s = 0.5gt^2$ to calculate? This is the crux of the problem because for simple ordinary differential equations, formulas like $s = 0.5gt^2$ are called analytical solutions. However, not all ordinary differential equations have analytical solutions. Therefore, for the vast

majority of cases, we actually solve them through iteration. These ordinary or partial differential equations only tell us the rules of iteration. For example, the equation set

$$\frac{dv}{dt} = \frac{F}{m};$$

$$\frac{dS}{dt} = v;$$

can be translated into instructing the computer to calculate

$$V_{n+1} = V_n + \frac{F}{m} * \Delta t;$$

$$S_{n+1} = S_n + 0.5 * (V_{n+1} + V_n) * \Delta t;$$

For more complex ordinary differential equations, analytical solutions may not exist. Take the following equation set as an example:

$$\frac{dx}{dt} = c_1 y + c_2 z^2 + c_3;$$

$$\frac{dy}{dt} = \frac{c_4}{y} + c_5 z - c_6 e^x;$$

$$\frac{dz}{dt} = c_7 + \frac{c_8}{x^2} + c_9 z;$$

Where $c_1 - c_9$ are constants, and x, y, and z represent variables. In this equation set, the change of variables over time is influenced by multiple variables at the same time, making analytical solutions often nonexistent. However, numerical methods can still be used for solution. In fact, even for those with analytical solutions, the earliest derivation and proof were often completed through iteration. Therefore, breaking down steps and continuous accumulation is the fundamental way to solve mathematical equations. The current methods we use to solve differential equation sets, such as Mathcad or Matlab, are based on iterative thinking.

Thus, we can see that in a mathematical model, the most important thing is to accurately describe the relationship between variables and time. By fitting it to actual data, we can obtain these parameters or variable parameters. An accurate equation set with strong physical principles will add strong applicability and predictive capabilities to the mathematical model. It is worth noting that even unreasonable equations can achieve good fitting effects by increasing the number of parameters, which can enhance the flexibility and pliability of the model. However, this cannot be transferred to other data, leading to a serious compromise in the reliability of the model.

CHAPTER 2

MISCONCEPTIONS OF HERD IMMUNITY

"Dogmatism kills"
—Deng Xiaoping

There is no theory in the world that is always correct. All theories have a certain timeframe and scope of applicability. Just like the classic Newtonian mechanics cannot explain the motion of the microscopic world, the theory of herd immunity cannot be applied to COVID-19 infections. The concept of herd immunity originated in the 1920s and 1930s when people were searching for a way to control the spread of epidemics by forming immunity within a community. This theory was initially proposed by S.W. Balfour in 1923, who believed that if as many people as possible were infected with a virus and gained immunity, the spread could be effectively controlled.

In 1933, Friedrich Föderl and George S. Baehr published an article titled "The significance of age distribution in pneumococcus infection," which explored the influence of herd immunity on bacterial infections. This article introduced the concepts of "immune population" and "susceptible population," and pointed out that introducing only a small portion of the population could achieve herd immunity for the entire community.

In 1940, Donald Henderson and other researchers proposed the concept of "basic reproductive number" based on their studies of other diseases. This concept was used to describe the number of people an infected individual can transmit the disease to in the absence of herd immunity. If this number

is below 1, the virus will die out within the population. However, if the number is higher than 1, the disease will spread extensively.

In the 1960s and 1970s, the theory of herd immunity began to be used to assess the effectiveness of vaccines. Researchers analyzed historical data on various diseases and established predictive models based on the principles of herd immunity to determine the optimal vaccine coverage rate. This led to successful large-scale vaccination programs, such as the eradication of measles and the action plan against poliomyelitis. In conclusion, the theory of herd immunity is a principle and tool for achieving herd immunity, which is of great significance in controlling the spread of infectious diseases and promoting vaccination.

For most of us, we were only exposed to this theory after the COVID-19 pandemic, when infectious disease experts hoped to promote vaccination through this kind of publicity. Now, let's delve into the theory of herd immunity in more detail. The premise of this theory is that when people are infected or vaccinated, they produce antibodies in their bodies, which protect them from secondary infections. It is obvious that if everyone in a population is vaccinated or infected and develops antibodies, secondary infections can be avoided. Here, we must explain the concept of the basic reproductive number, R_0. In simple terms, it represents the number of people an infected individual can transmit the virus to in a population without herd immunity. The higher the value, the more contagious the virus is. Mathematicians have theoretically deduced that when the value of R_0 is less than 1, the total number of infections will decrease over time. This is easy to understand: when one infected person spreads the infection to fewer than one potential case, the number of infections naturally decreases over time. Therefore, they believe that controlling the spread of infectious diseases only requires keeping R_0 below 1. For example, if the initial R_0 value is 3, meaning one person can infect three others, when two of those three

individuals are vaccinated, one person can only transmit to one other person. At this point, the R_0 value equals 1. In other words, for a virus with a transmission coefficient of R0, herd immunity can be achieved as long as the vaccination rate reaches $1-1/R_0$. What's even more interesting is that during the early stages of the COVID-19 outbreak, around early 2020, many simulation studies found that the virus seemed to predominantly infect the elderly population rather than the entire population. Based on this observation, they calculated that the susceptible population was much smaller than the total population, and the threshold for achieving herd immunity, as calculated based on this theory, was far below $1-1/R_0$.

However, it has been proven that herd immunity theory is clearly ineffective in the context of COVID-19 infection. This mistake is reflected in two aspects: First, even in the case of natural infection, the proportion of infected people in the population is significantly higher than $1-1/R_0$. Second, and most fatally, the protective antibodies produced are decaying, coupled with the effect of virus mutations, leading to multiple re-infections of COVID-19, similar to the flu. Many academic authorities, including renowned respiratory experts Wenhong Zhang and Nanshan Zhong in China, firmly believed in herd immunity theory in the early days of vaccine distribution in 2021. They believed that vaccination would completely eradicate COVID-19, just as humans did with smallpox. According to the theory of herd immunity, they even calculated that an immunization rate of around 80% could prevent the outbreak of the pandemic entirely. Looking back, this theory is absurd. However, we cannot deny its contribution. Driven by this theory, many people vaccinated themselves, which played a positive role in protecting their health.

Many scholars had discovered the errors of herd immunity theory as early as the end of 2020, when the COVID-19 virus had not undergone significant mutations, and its transmission coefficient had not undergone a sharp leap.

In many reports, Iranian researchers found that in some provinces, infection rates had already surpassed 80%, and in some areas of the Amazon basin, large-scale secondary infections occurred even after the first wave of high infection rates. At the time, the authenticity of these reports was also questioned because, according to herd immunity theory, it was impossible for the infection rate of natural infection to exceed the threshold of $1\text{-}1/R_0$. In early 2021, many Western countries began large-scale vaccination efforts, and China's large-scale vaccination also initiated in April 2021. The earliest vaccines available were adenovirus vaccines from AstraZeneca, followed by mRNA vaccines from Pfizer. In Western countries with smaller populations such as Israel, high-level vaccination had already been completed in the spring of 2021. Unfortunately, in June 2021, the pandemic resurged in Israel, which proves that herd immunity theory does not hold up when it comes to COVID-19.

In this chapter, we discuss why herd immunity does not apply to COVID-19 from three perspectives. Firstly, we introduce a new individual-based model used to predict the development and changes in the epidemic. In this model, we discuss why early predictions of the COVID-19 epidemic resulted in significant deviations. We also explain why the natural infection rate for infectious diseases significantly exceeds $1\text{-}1/R_0$. Secondly, we focus on explaining why COVID-19 antibodies cannot provide lifelong immunity. This was discussed in more detail in Chapter 5, we just provide a brief introduction here. Thirdly, we predict future trends in the epidemic, and our predictions have proven to be highly accurate.

Predicting the epidemic has become a hot topic since the outbreak of COVID-19. Many scientists, including Chinese academician Nanshan Zhong and Nobel Prize winner Michael Levitt, were interested in epidemic prediction. At the time, there were mainly two schools of thought. The first was that COVID-19 would destroy humanity. If not contained, it would

spread at a terrifying speed. If a person can infect two others, then two people will quickly infect four, and it will eventually rapidly spread to every person with exponential speed. This caused a great panic in the early stage of the epidemic. In China, the whole country adopted home quarantine measures, and in severely affected areas such as Wuhan, extreme isolation measures continued until May 2020. We must point out that the virulence of the COVID-19 strain was indeed very strong at that time, and many young people showed serious symptoms and even died. This panic quickly spread worldwide as the virus spread. What impressed me the most was that even the most developed medical system in the United States such as New York state, experienced a shortage of respirators. Governor Cuomo argued with the federal government over his need for respirators. At the time, he said that according to the epidemic curve predicted by infectious disease experts, there could be an exponential increase in infections in New York, which would mean a large number of critically ill patients requiring many respirators, far exceeding the government's inventory. Of course, there was a different idea, which was that COVID-19 would naturally disappear like SARS did, or even if it didn't, it would disappear quickly. The proponent of this theory was Michael Levitt and others. We will discuss why these two different predictions and ideas emerged.

Before discussing these two predictions, we need to introduce the classic mathematical model for infectious diseases. In 1927, Kermack and Mckendrick proposed the *SIR* model for infectious diseases, and since then, research on infectious diseases has been supported by mathematical models. *S* stands for susceptible individuals, *I* stands for infected individuals, and *R* stands for recovered individuals. The relationship between these three elements follows the following mathematical formula:

$$\frac{dS}{dt} = -\alpha SI;$$

$$\frac{dI}{dt} = \alpha SI - \beta I;$$

$$\frac{dR}{dt} = \beta I;$$

In the realm of scientific enlightenment, let us delve into the topic at hand in a manner befitting a piece of popular science literature. α pertains to the transmission coefficient of the virus, intricately intertwined with the acclaimed R_0 value, while β represents the recovery rate of the infected individuals.

The correlation between these components is remarkably straightforward. The rate at which susceptible individuals decline can be ascertained by multiplying α with the product of the susceptible population and the number of infected individuals. This phenomenon can readily be comprehended as infections stem from the contact between susceptible and infected individuals. Similarly, the fluctuation in the number of infected individuals can be delineated as the aforementioned value subtracted by the number of infected individuals transitioning into the realm of recovery. In turn, the emergence of recovered individuals stems from the transformation of infected individuals, encapsulated by the term βI.

By employing this mathematical model, we can evoke an intuitive sense and discern that the rate of infection will gradually intensify, only to subsequently diminish rapidly. Why does this intricate pattern arise, you may ask? The answer lies in the early stages of an epidemic where the susceptible population, signified by the variable S, remains relatively substantial, thereby impeding significant reduction. Consequently, infected population I continues to surge. Consequently, the term αSI, representing the number of individuals transforming into infected status, undergoes

gradual augmentation. Simultaneously, as the number of infected individuals escalates, the pool of susceptible individuals concomitantly decreases, causing a reduction in S. This, in turn, engenders a diminishing numerical value for αSI, implying a continuous decline in newly infected individuals.

Employing this model necessitates determining several parameters, namely α, β, and the initial number of susceptible individuals denoted as S_0. Once these parameters have been unequivocally established, we can prognosticate the future trajectory of an epidemic based on their collective values. However, a conundrum arises when divergent datasets yield disparate numerical values for these three parameters, ultimately yielding two diametrically opposed conclusions. Utilizing data from the early stages of an outbreak often yields a notably precipitous growth trend, thereby yielding relatively large values for the initial susceptible population (S_0). Consequently, predictions regarding the apex of infection tend to skew towards an elevated outcome, much like the projections bestowed upon Governor Cuomo of New York. Conversely, if we were to employ data gleaned from the Wuhan outbreak or any dataset exhibiting a discernible decline, revised data fitting would yield diminutive values for S_0. Predictions formulated based on this parameter combination allude to the notion that an epidemic can be promptly controlled and eventually eradicated, thereby attributing the source of Michael Levitt's and Academician Nanshan Zhong's erroneous conjectures.

The impact of their predictions during the period when the Chinese epidemic was teetering towards annihilation in 2020 had an indelible effect on me. Michael Levitt and Nanshan Zhong whimsically posited that a similar scenario would manifest on a global scale. Therefore, as I have tirelessly emphasized, embracing a simplistic approach reliant solely on data fitting does not bestow accurate prognostications. A profound exploration of underlying causes becomes imperative; otherwise, we

inadvertently navigate a path akin to forecasting stock market trends. Even the most formidable fitting algorithms and a plethora of extensive data, regrettably, prove ineffectual in such scenarios.

So what causes this deviation in predictions? The root cause can be found in the mathematical model itself, which is the original sin of this predicament. This is something I have always emphasized - when a theory has problems at its core, any improvements or developments based on it are fundamentally flawed. We must recognize the errors in classical models.

On a mathematical and logical level, this model appears to be perfect. However, it overlooks a crucial characteristic: population dispersal. The dispersal of animal populations is similar to the diffusion of chemical substances; it is not an instantaneous process but rather a gradual one. Take, for example, the discharge of nuclear wastewater in Japan. This wastewater does not suddenly appear in the Atlantic Ocean; its concentration gradually increases in the surrounding region and then spreads to distant areas. The same applies to human mobility in society. In fact, many scholars have long recognized this phenomenon, known as heterogeneity in media existence. This leads to the failure of the aforementioned mathematical model.

To help everyone understand more easily, let me provide a simple example. If we have a population of 10,000 people, and on average, each person comes into contact with two others, when one person becomes infected with a virus, they will transmit the virus to their two contacts. In the classical mathematical model, when these two individuals become infected, they would collectively transmit the disease to $2 * 2 * (10{,}000-3) / 10{,}000 \approx 4$ people, which is very close to 4 individuals. However, in reality, these two individuals will only transmit the virus to two other people because one of the individuals they had close contact with has already been infected and was the initial carrier. This is similar to the diffusion of bacterial colonies in a petri dish; they do not randomly and infinitely spread but gradually

expand from the edges. This type of transmission barrier caused by spatial constraints is something classical mathematical models cannot consider. The classical model assumes that if a droplet of water falls into a large container, the concentration will instantly be uniform throughout. This spatial contact situation causes a rapid decline in the R_0 value during the spread of the epidemic. Therefore, the observed R_0 value, even for the same strain, can vary greatly depending on the timing of data collection and the local population's contact frequency.

Using early data often yields larger R_0 values. For example, in the example mentioned above, if the number of second-generation infections is divided by the number of first-generation infections, we would think that R_0 equals 2. However, if we divide the number of third-generation infections by the number of second-generation infections, we would think that R_0 equals 1. This results in a very slow infection process within the population. The steepness of this curve is much lower than the steepness predicted by the *SIR* model. In a model representing a population with a large number of individuals and a wide geographical distribution, sometimes due to spatial diffusion, an epidemic in a large region may experience fluctuations. For instance, the initial peak of the COVID-19 outbreak in the United States was contributed by the eastern region, but later on, there were fluctuations caused by the transmission of the epidemic from one area to another. If we fit the data using only the first turning point and the data before it, we might wrongly assume that the first turning point represents the peak of the epidemic and thus obtain inaccurate results. These erroneous results include an underestimated initial susceptible population (S_0) and an incorrect estimation of the subsequent declining trend based on these parameters.

To solve this problem, we have developed an agent-based continuous Markov chain model. The principle behind the model is quite simple actually - by calculating the probabilities of individual agents, we can

determine the overall probability of the population's disease development.

Our article provides a detailed description of the Markov chain model, but for now, let's use a simplified example to illustrate our approach. We will assume that there are no complex factors at play and that all infections have the same potential for transmission. We will also assume that immunity to the infection is uniform across all individuals and that there are no viral mutations or antibody decay effects over time. Finally, we will assume that the overall population size remains constant.

Table 2.1 shows the infection probabilities of three individuals - A, B, and C - over several infection cycles based on this model. Using the contact matrix in Table 2.1, we calculate the initial virus reproduction coefficient $R_0 = 1/3 \times (0 + 0.8 + 0.5 + 0.8 + 0 + 0.6 + 0.5 + 0.6) = 1.267$. If A is the first to fall ill, we can use Table 2.2 to calculate the infection probabilities of A, B, and C.

Let's take a closer look at how we calculate these probabilities. Since A and B have a contact frequency of 0.8, the probability of A transmitting the infection to B would be the product of the probability of A falling ill and their contact frequency - which equals 0.8 in this case. Similarly, we can obtain the probability of C developing the infection in the second generation as 0.5.

For the third generation, B's infection probability comes from three sources - the probability of A infecting B, the probability of C infecting B, and the probability of self-transmission. Specifically, the probability of A infecting B is (B's susceptibility)(the probability of A falling ill in the second generation)(AB's contact probability) = (1-0.8)*0*0.8 = 0; The probability of C infecting B is (B's susceptibility)(the probability of C falling ill in the second generation)(BC's contact probability) = (1-0.8)*0.5*0.6 = 0.06; The probability of self-transmission for B is (B's susceptibility)(the probability

of B falling ill in the second generation)(BB's contact probability) = (1-0.8)*0.8*0 = 0. Therefore, B's overall infection probability in the third generation is 0.06.

By summing up the probabilities of each individual over *N* generations, we can obtain the overall infection situation of the population - for instance, the number of patients in the first generation would be one, and in the second generation it would be 1.3 (0.8 + 0.5). However, this simple model does not take into account factors such as viral mutations and immunodeficiency caused by antibody decay, which can increase the risk of re-infection. Therefore, we have designed a more realistic model where the sum of infection probabilities over time periods is no longer equal to 1.

For example, in the third generation of the previous example, the infection risk for B was (1-0-0.8) = 0.2. In the new model, this could become (1-0*c_1-0.8*c_2), where c_1 and c_2 are coefficients less than 1. For instance, c_1 = 0.5 and c_2 = 0.8. In this case, the infection risk becomes 0.36. Interestingly, when designing these coefficients, we adjust them to be closer to 1 for generations that are closer together in time, and gradually smaller and closer to 0 as the generations become more distant. This aligns with the real-world infection situation, where early infections offer less protection against re-infection over time.

Interaction Matrix	A	B	C
A	0	0.8	0.5
B	0.8	0	0.6
C	0.5	0.6	0

Table 2.1: Interaction frequency matrix among three individuals.

	A	B	C
1st generation	1	0	0
2nd generation	0	0.8	0.5
3rd generation	0	0.06	0.24
4th generation	0	0.02016	0.00936

Table 2.2: Infection probability of three individuals at different time points.

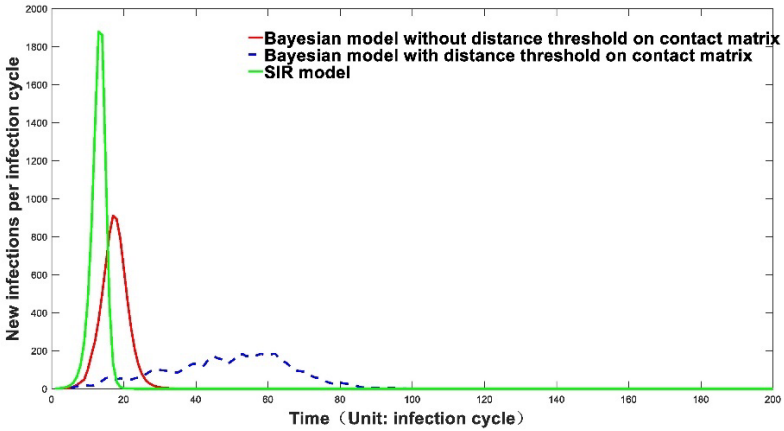

Figure 2.1: Epidemic trend predicted by three different models.

Before we present the simulation results of complex models, let's take a look at how simple models can challenge traditional understanding. Figure 2.1 shows that the early upward trend of the infection curve predicted by the SIR model is steep, while the upward trend of the Markov chain model considering population contact is relatively gentle. Our Markov chain model can be divided into two types. The first is a constrained model that considers population contact distance. The contact probability between individuals is inversely proportional to the fourth power of the distance between them.