COVID-19 VIRUS, MATH MODELS, AND MORE

DR. KAMAL LODAYA (RETD. INSTITUTE OF MATHEMATICAL SCIENCES, CHENNAI), DR. VANDANA REVATHI VENKATESWARAN (MAX PLANCK INSTITUTE FOR EVOLUTIONARY BIOLOGY, GERMANY), DR. N. D. HARI DASS (TIFR, HYDERABAD), DR. SHANTALA HARI DASS (INDIABIOSCIENCE)

Contents

1. Questions

1.0.1. *Germs, microbes, bacteria, viruses, what are they? Microbes* are tiny organisms which can only be seen through a microscope. They were seen by Anton van Leeuwenhoek in the 17th century. *Germ* is a common word for a microbe. *Protists (fungi, algae, protozoa), bacteria and viruses* are different kinds of microbes. Well, mushrooms are fungi which can be seen very well with the naked eye, but generally speaking the idea of a microbe being very small works.

1.0.2. *Why do they cause diseases in humans?* Many microbes are beneficial to humans. Bacteria are notable in helping our digestion. But they can also take away our nutrition for their own benefit.

Most of life seems to be organized in *cells*. Cells *replicate* to increase their number. Outside a living cell, viruses are inert and seem lifeless. When they enter a cell, they take over its material and start growing and replicating. Because they take over a cell's machinery, viruses are usually a problem. But if we want to cure ourselves of diseases which damage our cell machinery, viruses can be used as medical agents.

1.0.3. *Which diseases do we get from bacteria and viruses?* Cholera, tuberculosis, pneumonia, typhoid, leprosy are bacterial diseases. Flu (influenza), the common cold, rabies, Sars and Covid-19 are viral diseases. The common cold can be caused by two kinds of viruses: rhinoviruses and coronaviruses (which are different from the Sars variety).

1.0.4. *Apart from physical injury and the infections related to them, and what we get through food, are there any other diseases we get?* There are many non-infectious diseases, such as heart diseases and cancers.

1.0.5. *Is it true that hundreds of different bacteria have humans as hosts?* Yes. For every human cell in our body there are about 10 microbes, most of them harmless or even useful to us. A very small number are *pathogens* which make us ill. These microbes seem to play an important role in training our *immune systems* to recognize which microbes are friends and which are foes. Babies acquire them by age 3. Why do we humans outsource this function to microbes? Perhaps because they respond fast to adapt to changes in the environment.

1.0.6. *Which microbes affecting us have come to us from animals?* Rabies, Sars-Cov, Sars-Cov-2 (which causes Covid-19), Nipah and Ebola viruses have come from bats. (Rabies comes to us via dogs.) Turtles, snakes, frogs carry Salmonella viruses. Plasmodium, the microbe causing malaria, comes from mosquitoes. The dengue virus comes from mosquitoes.

1.0.7. *Why don't they affect bats?* Bats seem to have some of the strongest immune systems amongst animals. Many viruses which affect our immune systems do not seem to bother bats.

1.0.8. *How do they come from bats or whichever host to us?* There are various theories about an intermediate species for Sars-Cov-2 (which causes Covid-19), first it was snakes and later it was pangolins (a kind of ant-eater). We are not sure. For Sars-Cov we know that the intermediate species was civet cats. For swine flu the intermediate species was pigs.

1.0.9. *Does the intermediate species get affected as well?* Yes, they get disease as well. A mosquito with plasmodium is ill.

1.0.10. *How did this spread from bats to humans?* Coronavirus family is a huge family; just like the cat family where we have domestic cats, lions, tigers, etc. Some coronaviruses affect only other animals, that is, they are not harmful (pathogenic) for humans. One coronavirus, Sars-Cov-2 which causes Covid-19, has *mutated*—changed over generations—in a way that it can attach to humans as well, and survive.

Recent research suggests that even after attaching itself to humans, the Sars-Cov-2 virus might initially have not shown any effect on humans (it would have been *non-virulent* or *non-pathogenic*). This would later have developed virulence and turned into a pathogenic version, thereby affecting us. This happened earlier with the virus that caused the Ebola disease. So a recurrence of this outbreak is very unlikely. On the other hand, for the Sars virus, the pathogenic version developed in animals and then came to humans.

1.0.11. *Do humans also change?* Humans also change over generations. Over the last few centuries, Europeans have become more resistant to tuberculosis. Many earlier European novels, a famous one being Thomas Mann's *The magic mountain*, talk about *sanatoria* where Europeans went to recover from tuberculosis. Many West Africans have a type of red blood cell which gives them protection against malaria (it does give them anaemia, so it is not an unmixed blessing). Bangladesh (formerly part of the Bengal province of British India) has had a terrible history of malaria. It is known that people with blood group O are more susceptible to malaria. Now Bangladesh has the lowest frequency of O-group people in the world. The introduction of farming (that happened around 12,000 years ago) brought a large amount of disease to humans. About a 10,00 years ago, domestication of animals was introduced in humans, and that has also brought about a changed in humans; for example, genetic changes in humans allowed them to drink milk. A dental disease called "caries" was much more prevalent back then. The way our teeth are structured in the palate has changed since then.

1.0.12. *Do diseases get transmitted more easily in cities?* Depends on the disease. Diseases like dengue can be transmitted easily in rural areas and smaller towns too because it depends on water contamination. Diseases like dengue can also spread more easily in both cities and rural areas. However, diseases like Covid-19 spread through human contact. Therefore, there are more chances for it to spread in crowded areas and cities. If too much human physical contact happens in rural areas (for example due to festivals), then it will spread easily there as well.

1.0.13. *Do road, rail and air travel contribute to spread of disease?* Yes, they do if the disease is a contagious one with a sufficiently high transmission rate from person to person, and Covid-19 is one such disease. Data has shown that in this highly globalized and interconnected world, Covid-19 has spread through travel, and air travel has spread it to different continents. Refer to the blueprint diseases classified by WHO http://origin.who.int/blueprint/en/

1.0.14. *Do mobile phones help transmit disease?* It is not about phones. The issue is with our hands. When a virus carrier touches anything with their hands, there are high chances of disease spread. This can be stopped if we wash our hands with soap and water for 20 seconds before and after any contact (through phones, door handles, pens, etc) with other humans.

1.0.15. *Do wireless waves communicate disease?* No.

1.0.16. *Are couriers and delivery persons more subject to getting disease?* This can happen to anyone who is coming in contact with any affected (symptomatic or asymptomatic) person. This is not limited to domestic helpers, milkmen or courier delivery persons. This is why we need to reduce contact with others, and wash our hands with soap frequently.

1.0.17. *Does visiting crowded places contribute to disease?* Again, it depends on the disease. Many contagious diseases like Covid-19 spread through human physical contact. So yes, visiting crowded places during a contagious disease pandemic contributes to high levels of disease transmission from one person to another.

1.0.18. *Is Sars-Cov-2 natural or bioengineered?* Natural, according to recent research. Evolutionary genetics and biologists from the USA, led by Dr. Andersen and his team looked into this question scientifically, not just as a supposition. Their research has been recently reported in the journal *Nature Medicine* and showed that this virus is not bioengineering. It has to be a naturally evolved one.

They studied the genetic sequence of Sars-Cov-2. This was done to develop counter measures like vaccines, drugs, antibodies, etc. Its genetic sequence was released within few weeks after the virus showed that it can cause serious damage to humans. It was made public by China. The genetic material was made available to researchers across the world to study and develop drugs and vaccines. Dr.Andersen and his team compared this gene sequence with gene sequences of seven other coronaviruses that can affect humans and other animals (particularly bats and pangolins). They found two important results:

- Sars-Cov-2 was the least virulent one of the eight sequences. So this is useless for bio-war purposes. If one really wanted to create a bio-weapon, they would have created a highly virulent one. A laboratory manipulation of Sars-Cov-2 is analogous to forging a better knife. But the Sars-Cov-2 virus was much less efficient. If it was engineered, then it is like making a blunt knife.
- Imagine a lock and key situation by which virus receptors attach themselves to human receptors. The binding port ('key') of Sars-Cov-2 fits well with the human receptors ('lock'), and this kind of 'perfection' is achieved by natural selection and cannot be man-made. Despite the efficient binding, the interaction is not optimal. If a human created this virus, they would have created something that is optimal! Other coronavirus genes that were compared are much more optimal than the Sars-Cov-2 sequence. So this cannot be a purposefully manipulated virus. It has naturally evolved.

1.0.19. *Are there other coronaviruses besides Sars-Cov-2? Do they also cause disease in humans?* Yes, in fact the common cold can be caused by four kinds of coronaviruses. Sars is a respiratory disease caused by a coronavirus very similar to Sars-Cov-2.

1.0.20. *How did it spread so fast?* That's a big question. What we know is the virus spreads through *droplets* ejected by a patient during a cough or a sneeze. If the droplet falls on some surface, such as a doorknob or a table, it can survive for several hours. It can then move on to another person's hands, and when the person touches their face, to the nose, mouth or eyes. It can then enter the person's respiratory system and grow and replicate there. The swine flu virus H1N1 seems to have spread even faster.

1.0.21. *What makes Sars-Cov-2 so dangerous to humans?* Another big question we don't know the answer to. Compared to the Sars coronavirus which it resembles in its makeup, it has some structural differences. Perhaps these differences make it easier to move from human to human, compared to Sars. One possibility is that Sars-Cov-2 uses a site similar to a human enzyme called *furin*, but this is not sure. The Sars virus was more deadly, it killed about 1 person in every 10 it infected. Totally about 8000 people were infected and 800 died. At present it looks like Sars-Cov-2 kills about 1 person in every 40 it infects. On 31 March 2020 8 lakh people are infected and about 40000 have died. The swine flu virus H1N1 killed fewer people. It infected about 60 lakh people in 2009, about 20000 died. In India 34000 were infected and 2000 died.

1.0.22. *Why is this Covid-19 pandemic so huge?* We live in an interconnected world. Humans are meeting each other, trading, travelling and holidaying all over the world. Major tourist sites like the Egyptian pyramids and the Taj Mahal attract millions of visitors every year. Business is more globalized, today one can easily order and buy things made in other countries which are delivered to us.

2. Epidemiology of the Covid-19 disease

2.0.1. *Transmission and infection pattern of the virus.* Dengue is transmitted by mosquitoes. But this virus goes from one human to another. The coronavirus Sars-Cov-2 is in the droplets that come out from our nose. Sneezing may expel large quantity of the virus to the environment. The droplets drop down and settle on various objects such as door handles. Over time the virus loses its material but it seems to survive quite a few hours on hard plastic and steel surfaces. When we touch them the virus spreads to us. Further, when a person infected with the virus sneezes or coughs without covering the mouth, it may directly come to us if we are standing less than 1 meter away.

For around 14 days the virus can reside in a person it infects. After that it cannot survive in the body of that person because that person would develop *antibodies*, which are proteins made by our immune system to identify and help destroy the virus. Once a person is infected, generally she or he cannot be infected again, at least for years to come. Then how does the virus survive? It survives by spreading to another person. It needs a new host.

Each virus has a specific infection pattern. An infected person transmits the virus that is causing Covid-19 to another 2.6 persons, on an average, during the 14 day period. After that period she/he can neither host not transmit the virus.

2.0.2. *Exponential growth.* Now imagine an uncontrolled situation. One person is infected. That person transmits the virus to 2.6, let us say 2 people. Then these two can transmit to two each. Therefore in total 4+1=5 are infected. The four persons transmit to 8, eight to 16 and so on. In total, by 16th transmission the virus would have spread to 65535 people. This growth rate is called *exponential*.

All these 16 chains of transmission would happen in a matter of few weeks. Now on the 17th transmission the newly infected number will be, hold your breath, 1,31,070! Not all infected will exhibit extreme ill health. Only about 4.7% develop serious health conditions requiring medical attention. The value 4.7% of 1,31,070 is 6160. This means the new patients who need medical attention will be 6160! On that day alone, the patient number will double. If this continues to increase, one

need not tell, the medical facilities will be suffocated and there will be no room for providing treatment to all.

Data and studies have shown that this does exponentially rise each day and with every contact. This exponential growth has to be stopped. The exponential growth curve has to become 'flattened' or, in the best case, start dropping. Data from China and South Korea show how *physical distancing* drastically helps for this purpose. We need to 'break the chain' of the virus spreading. One of the important ways one can achieve this is by physical distancing. Large number of people should not make direct physical contact with other. This can be achieved by staying indoors.

2.0.3. *Lockdown of cities and curfew measures are important. Stay-at-home recommendations* are in order to reduce transmission, to avoid an increased number of unwell cases, and thereby decrease the pressure on the health system. If more and more cases turn up in hospitals, our healthcare system would crumble. So we need to reduce the number of cases per day, and this can be done by not allowing the virus to jump from one human to another. This is achieved if people stay at home, avoid mass gatherings and close contacts with people, especially known high-risk groups.

Lockdowns are where people are advised to stay and work from home as far as possible, with many facilities and offices closed. Temporary lockdowns for a day may not be adequate, yet, anything is better than none. The number of new cases can stop in high numbers (remember, we are talking about exponentially increasing values per day). However, just one day is not enough. What if a Covid-19 carrier (symptomatic or asymptomatic) steps out of their house the day after the lockdown, touches others' hands, sneezes and infects more people? So more such lockdown measures are required.

Curfew is a mandatory lockdown, for example, police may be asked to prevent people from coming out on roads.

2.0.4. *What is the difference between physical distancing, isolation, and quarantine?* All these measures are done to separate potentially or actually sick persons from healthy persons.

Physical distancing is adopted by two people who may have no symptoms. For example, when you cross someone on the street, you maintain a 1 meter distance, just in case the other person was infected, perhaps without even knowing it.

Quarantine is adopted for people who are healthy but suspected high-risk cases of Covid-19, for example if they may have had contact with people who might be

infected. If disease develops, even during asymptomatic or subclinical phases of the disease in an infected person, the virus can be transmitted to healthy persons, and this had to be avoided. During March 2020, all people who came by international flights to India were put in mandatory quarantine facilities if thermal scanners at the airport detected a high body temperature, others—presumably healthy—were asked to voluntarily quarantine themselves at home, even from family members. During the quarantining period (typically 14 days), they had to monitor themselves for Covid-19 symptoms. If symptoms were detected, a test would be carried out.

Isolation is adopted when we have a confirmed case of Covid-19. Moderate or severe cases are hospitalised to provide care; mild cases have to isolate themselves at home just as in self-quarantine, but this goes on until the patient recovers. However, if widespread *local transmission* occurs from persons having no travel history, then a blanket recommendation for individuals with symptoms to isolate at home may be given.

2.0.5. *We need to work to survive. How can the lockdown feed us?* Yes, not everyone is privileged to be able to work from home. But students can remain in their rooms and come out only when necessary. All physical classes are cancelled although some continue online.

Physical meetings and social gatherings all come to a temporary halt; however important they are.

Young citizens may not be affected so much, but they might be asymptotic carriers and can go affecting old people, and other people suffering from illnesses.

If you are a working individual, ask for your employer to give you paid leave for his/her own safety too! If they do not, then follow all the safety procedures described by the WHO. If your work allows you to take emergency or paid medical holiday, now is the time to do so!

If you have or know anyone who has cooks or domestic helpers coming to their homes, ensure they get paid leave.

Buy goods from small shopkeepers and vegetables from small vendors.

2.0.6. *Why is universal healthcare important in situations like these?* We are social animals, and we live in herds. One person's health can easily affect another. Doesn't matter if you are (or around) rich or poor. Pathogens affect all humans equally, and only by providing equal healthcare opportunities to each and everyone, can we save our entire herd.

3. The Covid-19 pandemic

3.0.1. *When and where was Sars-Cov-2 first seen?* The first patient to be diagnosed with Covid-19 had symptoms of the disease from December 1, 2019, in China. The Sars-Cov-2 virus was seen in medical labs in China in December. The older Sars-Cov was first found in November 2002 in China.

3.0.2. *Did it affect the Chinese because they eat many different meats?* It is possible that the Sars-Cov-2 virus may have jumped species at a meat market in Wuhan, China, because several early cases in December 2019 arose there. But this is no longer sure. Cooking vegetables and meat well destroys microbes. Many people are *allergic* to some foodgrains, eating these grains makes them ill.

3.0.3. *How did it affect China?* Covid-19 spread very fast in the Hubei province of China where the city of Wuhan (with 1 crore people) is situated on the river Yangtze. One reason may be that to celebrate the Chinese new year, many people from elsewhere in Hubei working in Wuhan went to their homes in January. People from Wuhan also went to other places in China taking the virus with them. Since it takes between a week to two weeks for disease symptoms to show, at first the seriousness of the disease was not recognized. The virus genome was mapped at the Wuhan Institute of Virology. The first patient to die was on 9 January 2020. Some doctors saw the similarity with the Sars epidemic, which had affected China in 2002, and alerted their friends. Dr Li Wenliang later died of the disease. Alarmed by its spread, the Chinese government responded with a *lockdown*, cutting off Wuhan and Hubei province from the rest of the country by stopping road, rail and air transport. Over the next month, the virus had spread to all provinces of China, and to many countries (including the United States) from where many people travel to China. Neighbouring South Korea was affected and it is suspected that North Korea has also been badly affected.

3.0.4. *How did it affect the rest of the world?* The World Health Organization first thought the disease could be localized in China. Some countries quickly cancelled flights to China, many did not. Italy cancelled flights to China, but the disease spread quickly there, it appears from just a few visitors from China, and today there are more deaths in Italy than in China. From Italy the disease spread to all of Europe and many other countries. Iran cancelled flights late, meanwhile the disease spread quickly there and from there to other West Asian countries. On 11 March 2020, the WHO declared the disease a pandemic. The US president made several statements decrying the seriousness of the disease. On 20 March 2020, two senators from his party who were on a government committee regarding the pandemic were accused

of *insider trading*: they were found to have made several million dollars by selling off shares early at a higher price before the markets crashed due to uncertainty. On 31 March 2020, the US had nearly 1.5 lakh cases of coronavirus.

3.0.5. *How did it affect India?* On 30 January 2020, a student from Kerala studying in Wuhan became the first Indian case. After that cases came from IT people coming from Europe and the US to Bangalore and Hyderabad, Italian tourists visiting Rajasthan and Indian tourists returning from the Middle East, especially via Dubai. The first person to die in India was on 12 March 2020, a pilgrim who came back from the Haj in Saudi Arabia to Karnataka. Because of a deliberate strategy of not testing for the disease and using curfews to avoid burdening India's hospital infrastructure, the number of known cases were around 1400 on 31 March 2020, around 35 deaths were reported. On 22 March 2020, India had a one-day janta curfew. Unfortunately it was coupled with an ill thought out show of national solidarity which led in several cities to more mingling of people. The curfew was extended for a week more in around 80 districts that day and for three weeks all over India two days later. Some medical researchers estimated that local transmissions were happening but were underreported, and there may have been 10 times as many Indian cases on that date, as the disease spread silently through people.

3.0.6. *What are the current crucial steps to follow?* Right now, the crucial thing is to stop the virus from spreading. By doing this, we give it no grounds to further evolve. Currently, we do not have full information about the virus evolution. There is a hypothesis that the drastic effects seen in Italy could be because the virus has mutated while spreading. Thus, we need to give researchers time to study the virus evolution and also to find vaccines, antibodies, etc. For giving them that time, we need to slow down the spread as much as possible. We need to act fast to flatten the exponentially growing curve quickly. We must follow physical distancing, lockdown measures (as much as possible by each of us), washing hands, etc. The WHO has warned that lockdown measures are not enough, we must find the virus and kill it. Even if we find no ways of immunity against it, as in the case of the human immunodeficiency virus which causes Aids, we can control it from spreading (like we do with HIV). We have done it in the past with many such viruses. We can do it this time too, if we all cooperate.

4. Some background

In this section, we look at three examples from the world of medical science of how mathematical models are built, tested and used. All three examples come from a book by Judea Pearl and Dana Mackenzie, *The book of why* (Penguin, 2018).

Then we give a sketch of what a disease model looks like with today's computer programs.

4.1. **How averages work.** Data science is a big thing these days. We give here an example of how tricky it can be to arrive at conclusions based on data, and how new medicines must be very carefully tested. The ideas were first pointed out by statistician Edward Simpson in 1951.

Let us suppose you read about a promising new drug which seems to reduce the risk of a heart attack. Here is the (fictitious) data that you read:

	No drug, attack	No drug, no attack	Drug, attack	Drug, no attack
Female	1	19	3	17
Male	12	28	8	12
Total	13	47	11	49

To your surprise, you see that the drug does not work for males, and it does not work for females. Yet it seems to work for people as a whole!

Among the 60 women, 1 in 20 (5%) in the *control group*—those who did not take the drug—later had a heart attack, compared to 7.5% of those who did take the drug, so the drug carries a higher risk. Among the 60 men, 30% of the control group later had a heart attack, compared to 40% of those who took the drug, so the drug carries a higher risk. But the third line of the table tells us that among the control group, 22% had a heart attack, whereas 18% of those who took the drug did. So the drug seems to reduce the risk of heart attack in the population.

In numbers, $3/40 > 1/20$ and $8/20 > 12/40$, but

$$(3+8)/(40+20) < (1+12)/(20+40).$$

The problem lies with understanding *fractions*: it is *not* true that if $A/B > a/b$ and $C/D > c/d$ then $(A+C)/(B+D) > (a+c)/(b+d)$.

More intuitively, the table introduces a spurious variable: *gender*, which is unaffected by the drug. (Further, men are at greater risk of heart attacks than women in general.) Statisticians would say that in this example, gender confounds the connection between drugs and heart attacks.

Working with *averages*, because men and women are equally frequent in the population, the rate of heart attacks without the drug is 17.5% (the average of 5 and 30), and with the drug is 23.75% (the average of 7.5 and 40). So the drug is no good for the population as well.

4.2. **How nutrition works.** Biochemistry is the science that explores chemical processes related to living animals. We tell you a story about how general dietary advice can be wrong, for lack of a precise understanding of nutrition.

In 1747, Scottish sea captain James Lind showed that eating citrus fruits prevented sailors from developing scurvy, a dreaded disease of malnutrition. It is estimated to have killed about 20 lakh sailors from 1500 to 1800. But tragedies were to follow more than a century later: Arctic expeditions in 1875, in 1894 and the Antarctic expeditions of Robert Scott in 1903 and 1911 all suffered greatly from scurvy. All the five people in the 1911 expedition who reached the South Pole died, two of an illness which was most likely scurvy. One team member returned before reaching the Pole and he had a severe case of scurvy. Why?

Well, why did citrus fruits prevent scurvy? Since Lind's time, it was taken as obvious, without evidence, that the reason must be their *acidity*. So one might as well have taken Coca-Cola on an expedition (although it had not then been invented) and survived! What sailors did was to substitute West Indies limes for Spanish lemons since they were cheaper. Then they started "purifying" the lime juice by cooking it.

When the sailors on the 1875 Arctic expedition fell ill with scurvy despite taking lime juice, doctors discarded the theory that citrus fruits could prevent scurvy. Suspicion instead fell on tinned meat, since sailors who ate fresh meat did not develop scurvy. Scott's 1911 expedition carried pemmican (a specially preserved meat dish having a lot of fat, of the Cree Indians living in the Canadian north), chocolate, sugar, biscuits, butter and tea. No citrus fruits.

It was only in 1912 that Casimir Funk, a Polish biochemist, proposed "vitamines" and in 1930 that Albert Szent-Gyorgyi, a Hungarian biochemist, isolated ascorbic acid—today called *vitamin C*—as the particular acid whose absence led to scurvy and malnutrition. It turns out that West Indies limes have much less vitamin C than Spanish lemons (although they have as much acid), and cooking lime juice further breaks down the vitamin.

4.3. **How epidemiology works.** Epidemiology studies occurrence, spreading and control of diseases and other issues related to health. We tell you here a story about how epidemiology began in 1854 with the study of cholera by a London physician, Dr John Snow.

At that time, cholera was a terrifying disease. But disease-causing germs had not yet been seen in a microscope. It was thought that cholera was caused by *miasma*, or unhealthy air. This theory was supported by the fact that the epidemic hit poorer people harder, and they had worse sanitation.

Dr Snow was skeptical about the miasma theory. He knew that the symptoms (diarrhoea) appeared in the intestines, so he thought the body must be encountering the pathogens there. He connected the victims to water pumps from where they got their water supply.

There were two water supply companies: the Southwark and Vauxhall, and the Lambeth. Both got their supply from the river Thames. Snow knew that the SV drew its supply from London Bridge, which was downstream of where London's sewers emptied. The Lambeth, on the other hand, had moved its intake to upstream of the sewers.

When he analyzed his data, Snow found that the death rate of patients drawing their supply from the SV was 8 times higher than of those getting it from the Lambeth. But of course the miasma could be higher in the poorer districts, so this was not definitive.

Snow observed that even in districts served by both companies, with the water going to large houses and small, the death rate was still higher in those houses getting SV water. This eliminated miasma and poverty from consideration. (Today we might call this a randomized controlled experiment with three lakh people, conducted by the two companies unknowingly.) The link between water companies and cholera was indirect—certainly they were not trying to give cholera with their supply—and went through a new variable, *water purity*.

Snow printed a pamphlet at his own expense explaining this, where his conclusion was that if the Southwark and Vauxhall company had moved its intake point upstream, at least 1000 people's lives could have been saved. The pamphlet sold 56 copies. But it was the beginning of epidemiology.

4.4. **How disease models work.** Suppose we want to write a computer program which reads a database of numbers and computes their average. What is a mathematical model of such a program?

The program has various *states*: a starting state with 0 numbers seen so far, and the total of the numbers seen so far is also 0. Given a state where N numbers have been seen with total sum S, the program makes a *transition*: it reads the $N+1$'th number, say it is M, and adds M to the sum S. Now it is in a state where $N+1$ numbers have been seen with total sum $S+M$. When all the numbers have been read, say N in all with sum S, it makes another transition to an ending state where it prints the average S/N.

This is a very straightforward program, but any program can be seen as a system of states and transitions, the run of the program tracing a path from a starting state to an ending state.

Suppose we want to make a model of some real-life situation. Now things are not so mechanical, but we can again model it using states and transitions. For example I may be in a state "ready-to-lecture" in my class tomorrow. The next day there are many possibilities: I can transition to a state "deliver-lecture" in my class. Or I may be ill and I may transition to a state "cancel-lecture". Or there is an unexpected holiday and I remain in state "ready-to-lecture". To model these situations, we label every transition with the chance that it will happen. For example, 99% of the time I will transition to "ready-to-lecture". Maybe 0.9% of the time I will transition to "cancel-lecture". Maybe 0.1% of the time I will remain in state "ready-to-lecture". This is only a model, for example, the government may declare a lockdown tomorrow and I may transition to state "cancel-lecture" but I may not even have thought of that possibility. Such models are named after the Russian mathematician Andrei Alexeyevich Markov who studied them in 1906. (His son Andrei Andreyevich Markov was also a well-known Russian mathematician, sometimes there is confusion between the two.)

Suppose we want to make a model of how some disease like Covid-19 spreads. Again we use states and transitions. For example, a patient may be in a state "susceptible" and can transition to state "infected", then transition to state "recovered" or state "dead". Now we come up with a problem: suppose to begin with the entire population is in the susceptible state. What is the chance that a patient remains in this state? What is the chance that a patient gets infected? For an unknown disease we have no idea of the chances.

We begin with some data about how the disease behaves with patients, perhaps data about the first 100 patients in Wuhan. That gives us some percentages about the chances that they will move from one state to another. Now we have to see how these percentages can *change*, because as there are more infected people they can go about infecting more of the population. Modellers write equations for such changes in percentages, like we do in school algebra. The state-transition network now models a situation changing over time; sometimes this is called a *dynamical system*.

Now we use a computer program to run with "sample" patients over many very short time steps, possibly making transitions between different states. These runs can be put together—this is called *numerical integration*—to calculate percentages of the chances that sample patients will end up in different states, for example, the percentage of people who will recover and the percentage of people who will

die. This technique is named after the city of Monte Carlo in Monaco—famous for its gambling casinos—since it appears as though we are "gambling" with patients' lives, of course this is done only in the model for these "sample" patients! Which is why models are useful - we only 'gamble' with ''sample' computer/artificial patients, even if it is done logically. The knowledge we gain from such simulations can then be used for application to real-life.

Ironically, on 17 March 2020, Prince Albert II of Monaco became the first head of state to be infected by the coronavirus.