Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

Mathematical Biosciences and Engineering, x(x): xxx–xxx
DOI:
Received: October 2020
Accepted:
Published:

Original research

Predicting COVID-19 using past pandemics as a guide: how reliable were mathematical models then, and how reliable will they be now?

Christian Costris-Vas[1], Elissa J. Schwartz[2,] and Robert Smith?[3,*]

[1] Department of Mathematics, The University of Ottawa, 150 Louis-Pasteur Pvt, Ottawa, ON, Canada, K1N 6N5

[2] Department of Mathematics & Statistics and School of Biological Sciences, Washington State University, PO Box 643113, Pullman,WA, 99164-3113, USA

[3] Department of Mathematics and Faculty of Medicine, The University of Ottawa, 150 Louis-Pasteur Pvt, Ottawa, ON, Canada, K1N 6N5

* **Correspondence:** rsmith43@uottawa.ca; Tel: +1-613-562-5800 x3740; Fax: +1-613-562-5776.

Abstract: During the earliest stages of a pandemic, mathematical models are a tool that can be implemented quickly. However, such models are based on meagre data and limited biological understanding. We evaluate the accuracy of various models from recent pandemics (SARS, MERS and the 2009 H1N1 outbreak) as a guide to whether we can trust the early model predictions for COVID-19. We show that early models can have good predictive power for a disease's first wave, but they are less predictive of the possibility of a second wave or its strength. The models with the highest accuracy tended to include stochasticity, and models developed for a particular geographic region are often applicable in other regions. It follows that mathematical models developed early in a pandemic can be useful for long-term predictions, at least during the first wave, and they should include stochastic variations, to represent unknown characteristics inherent in the earliest stages of all pandemics.

Keywords: SARS; MERS; H1N1; COVID-19; reproduction number

1. Introduction

COVID-19 is a respiratory disease with flu-like symptoms that originated in the city of Wuhan, in the Hubei Province of China at the end of 2019 [1]. On January 22, 2020, there were 425 laboratory-confirmed cases; since then, it has spread throughout the world, intruding on almost every aspect of daily life and negatively affecting the world economy [2]. By June 10, there were 7.45 million cases in

215 countries. The virus is especially deadly to older adults and those with underlying health conditions [3].

Along with the onset of the epidemic, the scientific literature has soared with publications on COVID-19. [4]. Several early papers have used the initial cases to build models describing the transmission dynamics and epidemiological characteristics of the virus. Some of these include the basic reproduction number (R_0), defined as the average number of secondary infections caused by a single infectious individual in a wholly susceptible population, and the effective reproduction number (R_e), defined as the average number of secondary cases per infected individual in a population with both susceptibles and non-susceptibles [5]. Unlike the basic reproduction number, R_e takes into account interventions, behaviour changes and control measures [6].

Early studies on the spread of COVID-19 found R_0 values that have ranged from 2.0 to 6.47. Li *et al.* used a transmission model with renewal equations and the first 425 cases of infection to predict an R_0 of 2.2 (95% CI, 1.4 to 3.9) and a doubling time of 7.4 days [2]. Tang et al., using a compartment model with quarantine and isolation interventions — along with the numbers of reported daily cases in Wuhan and quarantined and released/cured individuals for a two-week period — and found an R_e of 6.47 [7]*

Imai *et al.* used data from the initial weeks of infection to estimate an average R_0 of 2.6 (uncertainty range 1.5–3.5). They used two key characteristics that were established from the SARS epidemic to model the potential epidemic trajectories: that the number of secondary infections from each infectious individual is highly variable, and that the average time between generations of infection is variable as well. They approximated the number of cases at the end of the study to be 4000 (uncertainty range 1000–9700) using international travel frequency [8]. Zhao *et al.* used a statistical exponential growth model and the epidemic curve of cases in mainland China from two weeks early in the outbreak to estimate mean R_0 values from 2.24 to 3.58. This estimate also used serial interval data from SARS and MERS. However, they acknowledged that the official diagnostic protocol released by the WHO on January 17, 2020, probably caused an increase in diagnosis and reporting of infections [9].

Wu *et al.* used stochastic Markov Chain Monte Carlo methods with data from the first month of infection to find an R_0 estimate of 2.68 (95% CI, 2.47–2.86). The study predicted an epidemic doubling time of 6.4 days and that 75,815 individuals (95% CI 37,304–130,330) were infected in Wuhan as of January 25th, 2020. Serial interval estimates were again based on previous studies of SARS [10]. Riou and Althaus (2020) used stochastic simulations of early outbreak trajectories to calculate an R_0 of 2.2. An important aspect they identified was SSEs (super-spreading events), in which a few individuals infect many susceptibles in a short span of time. Using data from January 18th, they modelled SSEs using negative-binomial offspring distributions. With the data available at the time, they projected that the virus exhibited a mostly homogenous distribution pattern; i.e., SSEs were unlikely, though still possible [11].

Here, we investigate the following questions: 1. What do the dynamics of a disease in its early stages tell us about long-term predictions? 2. How well do models formulated with data from one geographic region predict data in other geographic regions? 3. Do certain methodologies do better than others in their predictive capacity? We reviewed modelling studies from recent pandemics in order to

*Tang et al., along with some other authors, use the parameter R_c, which measures the average number of secondary infections in a population when interventions such as quarantine, isolation and closed borders have been enacted. We have changed notation to R_e throughout for consistency.

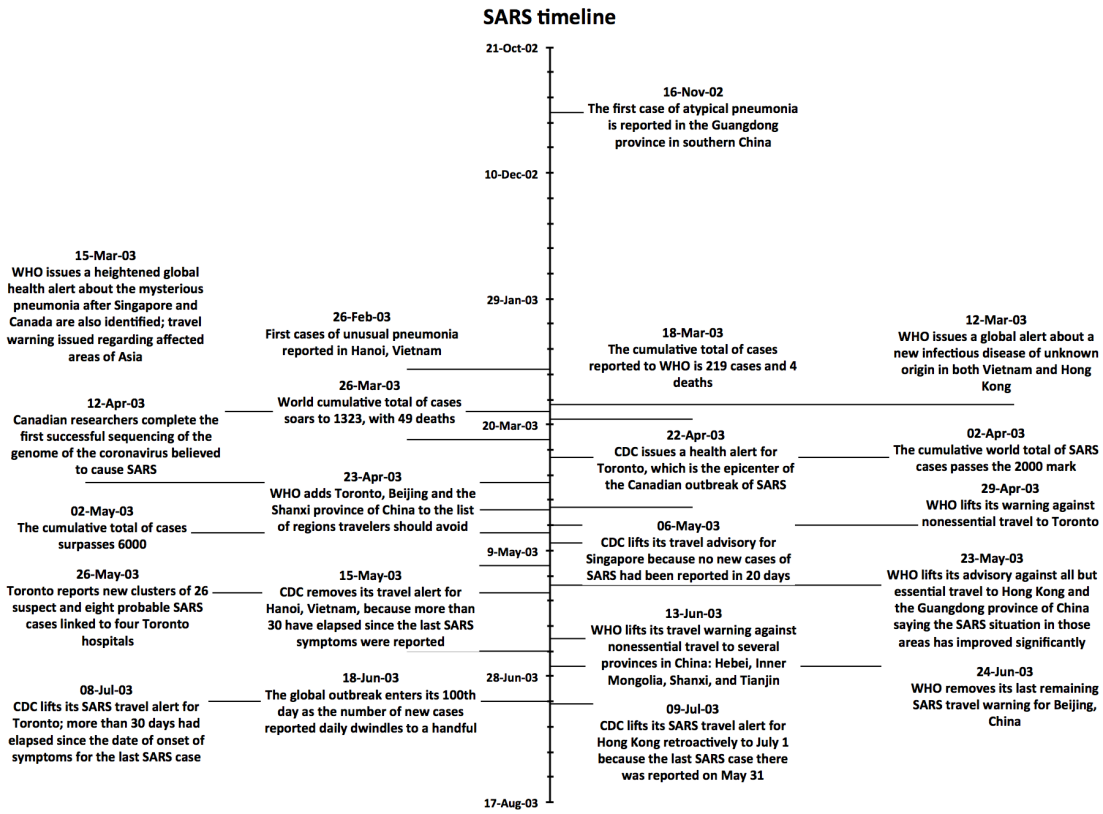

Figure 1. SARS timeline

assess how models that were created very early on in an outbreak compared to the eventual outcomes for those pandemics. We examined three major epidemics preceding COVID-19: the 2003 Severe Acute Respiratory Syndrome coronavirus outbreak (SARS-CoV), the 2012 Middle East Respiratory Syndrome-related coronavirus (MERS-CoV) and the 2009 H1N1 influenza pandemic (aka swine flu). We used studies performed using the first few weeks of an outbreak and compared them to the eventual results, in order to assess the reliability of early models. We also investigated which types of models produced the most accurate predictions. Finally, we determined scenarios in which early studies are likely to project the severity of a disease, those where they cannot, and potential pitfalls that could cause inaccuracies.

2. The 2003 SARS outbreak

SARS-CoV was a viral respiratory disease first seen in the Guangdong province of southern China in November 2002. It spread across continents to many of the largest cities in the world in an eight-month period until June 2003. The virus had spread to a total of 17 different countries, with the majority of cases in mainland China and Hong Kong. By July 2003, the virus had infected 8448 people, with a total of 774 deaths [12]. Secondary regions, which recorded fewer than 1000 cases, included Taiwan, Canada, Singapore, Vietnam, the US and the Philippines; the remaining countries saw fewer than 10 cases [13]. See Figure 1.

A limitation of these data is that government and health-agency statistics only list infections re-

ported to hospitals and health authorities; they do not include asymptomatic cases or cases that did not come forward for treatment (e.g., self-isolating individuals or those simply not seeking treatment). We restricted our scope to papers that used data specifically from the early weeks of the pandemic or before the disease was completely eliminated. Papers with sparse or incomplete data are of interest to see how well they predict metrics with these built-in limitations, as compared to post-epidemic publications.

We examined six papers studying different regions that used different methodologies for their models. Riley *et al.* analyzed data on the first 1512 cases in Hong Kong to determine a basic reproduction number of 2.7 at the start of the epidemic. They used a stochastic metapopulation compartmental model; this accounted for the high variability in transmission between parts of Hong Kong and changes caused by chance versus interventions. The populations of each district were divided into susceptible, latent, infectious, hospitalized, recovered and dead individuals; the latent and infectious categories were each divided into multiple categories to match disease progression. Despite the variability in the contact rate between districts and in the strength of SSEs, the output range for R_0 of 2.2 to 3.7 still suggested low to moderate transmissibility [14].

Chowell *et al.* estimated the reproduction number of the disease at 1.1–1.2. In this study, the authors used the cumulative number of SARS cases from two weeks in the beginning of the epidemic in Ontario, Hong Kong and Singapore and a compartmental model that accounted for known disease characteristics at the time. Points of uncertainty that were identified included asymptomatic infectiousness, levels of susceptibility among the higher susceptible class and lower susceptible class, the rate that the infected are diagnosed and the relative measure of reduced risk among diagnosed SARS cases. The authors discussed the possibility of 10% of some regions of the world becoming infected [15].

Lipsitch *et al.* analyzed the first 205 probable cases of SARS in Singapore with a stochastic transmission model to estimate an R_0 of 2.2–3.6, depending on the serial interval. The large variability in R_0 was attributed to the uncertainty in the distribution of secondary infections from each source case and the serial-interval distribution; the magnitude of SSEs can have a drastic effect on the number of secondary cases when present at the beginning of the outbreak. The frequency of SSEs is also difficult to predict using only the early disease data. Also unknown at the time was whether asymptomatic individuals were contagious, although the model was robust to the possibility. What it established, however, was that the few interventions that were being used were working [16].

Zhou & Yan used a Richards curve to fit early infection data from Singapore, Hong Kong and Beijing to find reproductive numbers of 2.7, 2.1 and 3.8, respectively [17]. They also made an attempt at predicting the maximum number of cases over the epidemic: 1748 for Hong Kong and 207 for Singapore; the actual numbers were 1755 for Hong Kong and 238 for Singapore [13]. The 2595 estimate for Beijing was further from the actual number of 5327. Zhou & Yan mentioned that one reason for Beijing's higher R_0 was their delay in exercising effective control measures. This highlighted a difficulty faced by all of the research in 2003: the transmission mechanism of the virus was not known at the time. Furthermore, timely and appropriate intervention on the part of health officials could not be assumed (as in the case of Beijing) [17].

Choi & Pak used Canadian data up to April 6th, 2003, and a simple exponential model with an R_0 of 1.5 to show that it could be possible to see up to 50,500 total cases without intervention by June 25th [18]. Lloyd-Smith *et al.* used Hong Kong and Singapore data prior to May 23rd to simulate the spread in a single hospital and show an R_0 of 1.2–2.0. This was thought to be a reasonable range given the data available at the time [19].

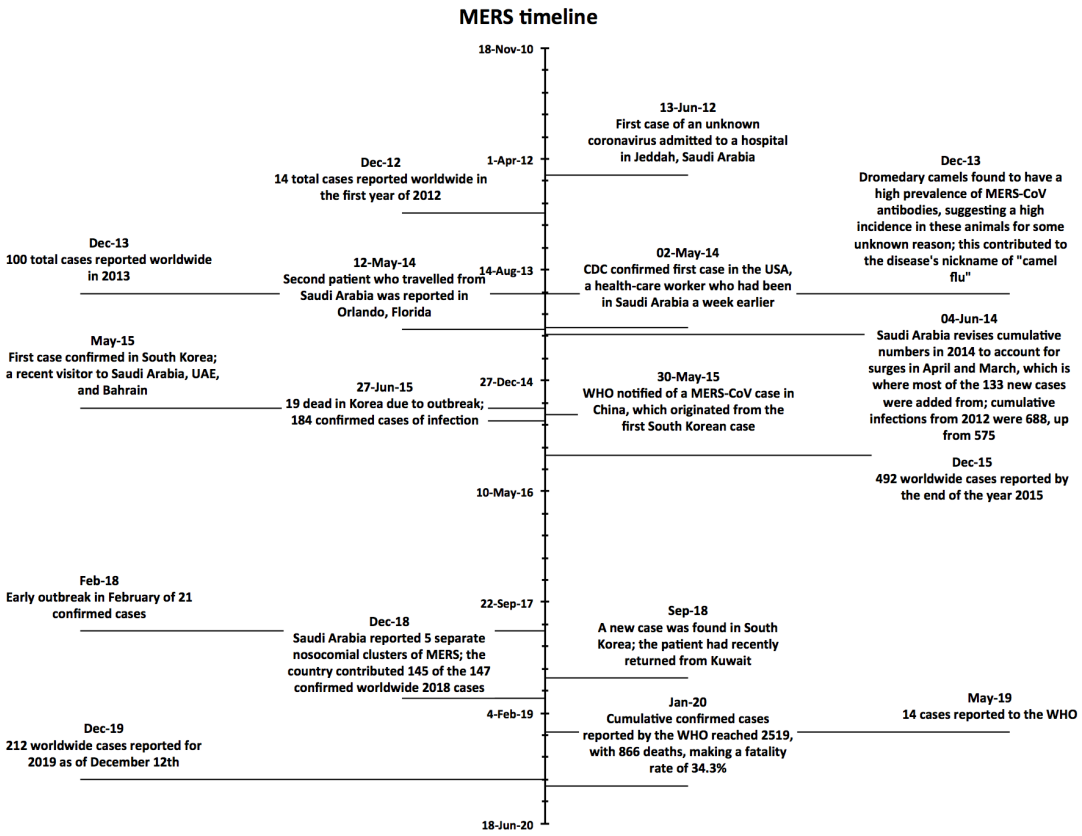

Figure 2. MERS timeline

3. MERS

Middle East Respiratory Syndrome coronavirus (MERS-CoV) is a respiratory disease similar to SARS and COVID-19 that has caused recurrent outbreaks in the Arabian Peninsula and other parts of Asia since 2012 [20]. See Figure 2. Symptoms include fever, cough, diarrhea and shortness of breath. It is suspected to have derived from bats and is found prominently in camels. The disease has been known to have a somewhat different transmission pattern to SARS; it typically has low transmissibility but occasionally sees temporary, rapid outbreaks originating in hospital settings [21, 22]. Its characteristic epidemic pattern can lead to different research questions compared to SARS; for example, could early rapid outbreaks inform the severity of later ones?

Cauchemez *et al.* used data up to August 8th, 2013 (111 total cases) to give an estimate of R_0 averaged across all detected clusters as 0.63. The majority of cases were in Saudi Arabia, followed by Jordan, the United Arab Emirates and a handful in some European countries. The doubling time was estimated to be 90 days. This study focused on clusters of non-resident travellers (non–hospital originating cases), and they acknowledged that it would make sense to see a reduced rate of infection because of inhomogenous mixing patterns [23].

Eifan *et al.* used data from the two-year period May 2013–May 2015 to assess Saudi Arabia's experience with the disease according to specific clusters. The country was divided into five different clusters based on their geographical separation; north, south, east, west and central. In the baseline

case, a sequential Bayesian method revealed a basic reproduction number in the range 0.85–0.97 in all five regions. However, results from a sensitivity analysis showed that the central and western regions achieved values between 1.08 and 1.12. The authors describe contact frequency, hospital procedures and population composition/density as being factors for the differences between regions. It is clear from their results that the virus is usually not sustainable to the point of being an epidemic threat [20].

A study by Breban *et al.* estimated R_0 to be 0.69 using only the first 64 (55 actually used in the model) lab-confirmed cases. The lower number here appears to be due to optimism about potential contact tracing and the isolation procedures that could be used during outbreaks. They concluded that MERS did not have pandemic potential, although they still highlighted the importance of enhanced surveillance and vigorous searches for the original animal hosts. While many studies used datasets with hundreds or thousands of data points, this study generated an R_0 in line with many other papers of the time using fewer than one hundred cases [24].

3.1. Super-spreading events

Both SARS and MERS nosocomial infections were characterized by early SSEs that briefly saw very high effective reproductive numbers, which then tapered off very quickly. This progression was recognized by Chowell *et al.* who used four different nosocomial outbreaks of SARS and MERS to show the effect of super-spreading events in hospital settings [21]. They focused on SARS outbreaks in Singapore and Toronto and MERS outbreaks in Al-Hasa, Saudia Arabia, and South Korea, creating transmission trees to quantify the average exposure time by type of individual (healthcare worker, patient, visitor, non-clinical staff) and were able to assign a reproduction number to each generation of the disease; the Generation 0 (R_0) of MERS in South Korea and SARS in Singapore had an overwhelming effect, with the overall numbers being estimated at 30 and 25, respectively [21]. The Generation 0 of MERS in Al-Hasa was estimated at 3; SARS in Toronto's had a Generation 0 of 1, which progressed to an epidemic with an R_e of 7 in Generation 1 and an R_e of 5–6 in Generation 2 before tapering off [21].

Hospitals were disproportionately large carriers at the outbreak of both diseases: healthcare workers must rotate, and family members visited patients [21]. A separate study by Chang estimated R_0 for the Korean case at 8.1 [25]. Despite being an order of magnitude higher than that calculated by Chowell et al., its value exceeding unity confirmed the epidemic threat.

4. The 2009 H1N1 epidemic

The influenza A (H1N1) virus that emerged in April 2009 spread worldwide and generated a pandemic that lasted until August 2010. The virus likely jumped species from influenza in pigs. Symptoms include nasal secretions, chills, fever and lowered appetite resulting from upper respiratory tract infections. Despite being a subtype of influenza virus A, seasonal flu vaccines offered almost no cross protection against the strain. Whereas typical seasonal influenza epidemics see the vast majority of deaths in individuals aged 65 or older, 80% of those caused by H1N1 were people under 65 [26, 27, 28]. See Figure 3.

Roberts & Nishiura used a renewal process to model the first two months of cases in New Zealand in a group of students returning from Mexico. The model incorporated imported cases alongside local growth and accounted for the infection-age distribution, determining an R_0 estimate of 1.25 (95% CI

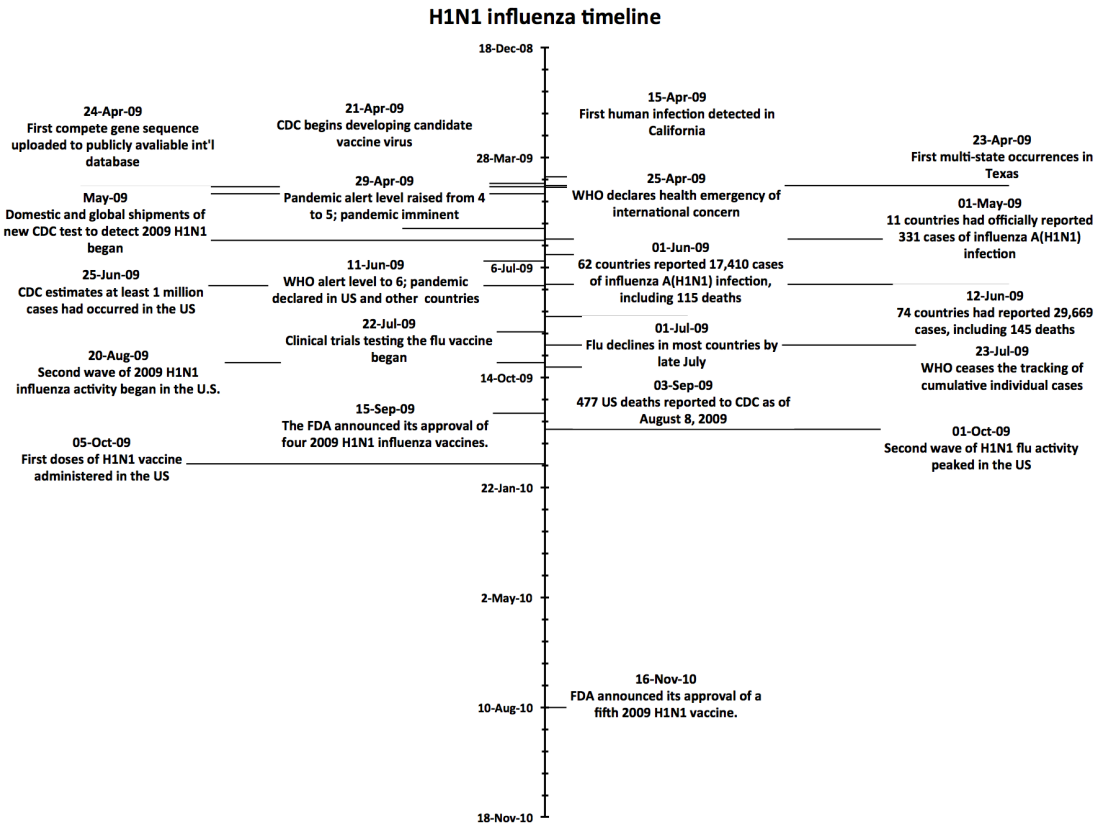

Figure 3. H1N1 timeline

1.07–1.47). The aim of the modelling was to show the effectiveness of different distributions for finding the best fit and to illustrate the importance of demographic stochasticity in modelling the early stages of the disease [29].

Using Mexican data, Fraser *et al.* predicted an R_0 of 1.4–1.6 in three different epidemiological analyses; a separate genetic analysis gave an estimate of 1.2. They used data on cases among travelers and back-calculation methods to estimate the total number of people in Mexico with the disease, using a cutoff date. This model used several different methods for calculating R_0, using 1) estimates for the cumulative number of infections based on the generation-time distributions of other seasonal viruses, 2) a Bayesian coalescent population genetic analysis and 3) the time series of reported disease onsets among confirmed cases. It also emphasized the importance that age-specificity played in fitting the model to the data accurately [30].

Nishiura *et al.* focused on age-specific estimates using Japanese data. They predicted R_0 values to be as high as 1.9–2.3. The study considered the 361 confirmed cases up to June 1, 2009. The authors suggested that the higher R_0 value was likely due to high contact rates among teenagers and suggested that minors could sustain transmission of the new influenza A (H1N1) virus among themselves. Of the 361 confirmed cases, 287 (79.5%) had been teenagers (10–19 years old) [31].

Estimates for the basic reproduction number found after the epidemic had run its course (post-2011) usually hovered around 1.3–1.7, with some studies for countries in the northern hemisphere describing values slightly above 2.0 [32, 33].

4.1. Waves

Pandemics typically come in waves, with some locations experiencing second and sometimes even third waves of the disease that were stronger than the initial one. While estimates can be derived for the spreading capability of the first waves using their specific data, these data are unable to predict the potency of subsequent waves or how likely they are to occur. Helferty *et al.* examined the effects of H1N1 in Canada and concluded that the second wave was substantially greater than the first. They reported that there were 4.8 times more hospital admissions, 4.0 times more ICU admissions and 4.6 times more deaths in the second wave than in the first [34]. In the UK, the second wave of H1N1 caused significantly higher population mortality than the first, with a rate of 5.4 vs. 1.6 deaths per million. Of the cases that qualified as being due to H1N1 over the whole pandemic, there were nearly four times as many deaths in the second wave as in the first (279 vs. 82) [35].

This disparity is reflected in the mathematical models. Using data in the UK through the pandemic, Dorigatti *et al.* found that, from an assumed R_e of 1.4 at the beginning of the outbreak, increased transmissibility grew R_e to 1.5 by the start of the third wave. Transmissibility was estimated to be between 67% and 121% higher than that found in Wave 1, depending on the model used and if prior immunity is considered [36]. Sharomi *et al.* built a model mirroring the data found in Manitoba, Canada; assuming the second wave for this region began in early October 2009, the reproduction number corresponding to this period was projected to be $R_e = 1.91$ [37] (the pandemic began in April 2009, but some sources suggest the second wave began in August 2009, depending on the region).

While a number of countries — including the US, UK, Mexico and Canada — experienced more than one meaningful wave of the virus, several countries, especially in the southern hemisphere, only experienced a single one [37, 38, 39]. Determining factors that predict single vs. multiple waves would be extremely helpful: what was about Canada, the US and the UK that encouraged a second, more potent, wave compared to other countries? Possible factors include the weather, population density, social pandemic procedures and the healthcare systems. Finding ways to predict when subsequent waves are likely to occur or what measures could prevent them would be enormously beneficial.

5. How Did We Do?

The SARS epidemic had mostly run its course by December 2003; there were only three confirmed cases and one probable case by the end of February 2004 [40]. Several studies reflected on the spread of the disease after the fact. Anderson *et al.* (published June 2004) determined the reproductive number to be between 2 and 3 [41], while Ruan et al suggested a range of 4–5 [42]. Many retrospective studies agreed on a value within 2–4, depending on the method used [43]. Several studies did accurately predict the strength of the epidemic, despite the short timeframe, limited available information and gaps in the data. Riley et al., Lipsitch et al., and Zhou & Yan had their estimates directly in the confirmed post-disease range [14, 16, 17], while others had their estimates slightly outside the range.

Accounting for and properly estimating government response is an important factor in modelling an outbreak. China's government and healthcare structure likely played a role in allowing SARS to persist beyond 3000–4000 total cases, as estimated by Zhou & Yan [17]. Estimates based on Toronto's or Singapore's healthcare systems were optimistic for a society of China's capabilities. This highlights the unpredictability of agents at the macro level versus the individual level. Also difficult to predict is the public's response to news of the virus and the magnitude of isolation once the pandemic is

underway. Furthermore, the capability of hospitals isn't always straightforward.

There have been few cases of MERS in recent years, most of which have been in Saudi Arabia. Apart from cases of SSE-triggered, high-R_0 outbreaks, recent MERS studies describe the reproduction number to be similar to that in papers that came out in the first few years after the disease's first case. Kucharski and Althaus (published in 2015) estimate R_0 to be 0.47 [44], which is in line with the earlier estimates from Cauchemez *et al.* (R_0 of 0.63) [23] and Breban *et al.* (R_0 of 0.69) [24]. In the cases of sudden, high-transmissibility events, the same trend of a high basic reproductive number at the outset (usually in the 1–7 range) with the effective reproductive number very quickly tapering off as individuals respond to the outbreak has been reported [45, 46, 47, 48].

The degree to which early epidemiological data was useful in predicting the 2009 H1N1 outbreak depends on the scale of the prediction; early studies did a good job of mapping the trajectory of the virus in its initial stage. This holds true when the epidemic passes in one continuous wave. The same is not always true in the case of multiple waves; while some subsequent waves maintained an R_0 similar to that of the early stages of the first [49, 50, 51], others increased their ability to transmit. Subsequent waves of a virus are not easy to predict, and neither is the increase in their strength.

Early disease models can thus do well in projecting the spread of a single wave. However, initial studies are not representative of the disease in later waves. Subsequent waves can lead to higher transmissibility, worse symptoms and increased mortality rates, as seen in Canada and the UK for the 2009 H1N1 outbreak.

6. Discussion

Mathematical modelling has the potential to make surprisingly accurate predictions in the early stages of an epidemic, despite the paucity of data. By comparing early models to what was later known about these outbreaks, we were able to answer our research questions. 1. The dynamics of a disease in its early stages can be used to make good predictions about the length and severity of a disease's first wave; these predictions are less good at characterising either the existence, timing or strength of subsequent waves. 2. Models formulated in one geographic region generally have good predictive power for other geographic regions in general, but do not always translate. 3. The methodologies that have the most success generally included stochastic effects. We thus recommend that future models include stochasticity, especially when modelling new outbreaks.

However, in addition to the success stories, lessons can be learned from models that were less successful in predicting past pandemics. For example, the nature of SARS transmission was not fully understood early in the outbreak; it was unclear at the time whether contaminated surfaces contributed meaningfully to the spread [15]. This uncertainty would have affected parameter estimates and model construction. Similar questions arose early in COVID-19, although it should be noted that a key difference between SARS and COVID-19 is that SARS was contained after its first wave [40].

Initial literature reports typically account for and describe individuals who have severe clinical infections; these are reported/admitted cases only. Asymptomatic cases that were not reported are either missed or need to be estimated. As such, retroactive studies have the benefit of more in-depth serological surveys that more accurately describe the number of infected individuals. The rate of infection during the asymptomatic period of the disease is often subject to uncertainty, especially in the absence of detailed tracing for reliable estimates. Some of the early MERS studies only looked

at community-acquired MERS infections and ignored nosocomial outbreaks [46]. Moreover, it is extremely difficult to predict where an SSE-level individual will be found and how much of an effect an SSE will have near the beginning of the outbreak.

It should be noted that diseases with low to moderate transmissibility display more uniform behavior and are easier to model and project; diseases with more volatile spread rates are less likely to display this uniformity. While diseases with lower reproductive numbers usually have less potential to spread in a devastating way, they can sometimes infect a lot of people; diseases with very high reproductive numbers have the possibility of infecting very many people but can sometimes only infect very few, depending on clusters.

Models with an explicit stochastic component seemed to perform well against compartment models that derived their R_0 values in more traditional ways. These models placed a greater focus on quantifying the effect of the super-spreading events and how their involvement at the start of the disease could be the difference between a large outbreak or not. For example, whereas Chowell *et al.* considered the role of super-spreaders in the parameters [15], Riley *et al.* tried to incorporate an estimate for the explicit number of individuals who were infected by the super-spreaders and who went on to propagate the disease [14]. Other factors included taking into account differing contact rates in different regions of the study; for example, dividing Hong Kong into higher and lower contact districts. Choi & Pak incorporated stochastics into a compartment-based model and only split these compartments up when necessary [18].

6.1. Lessons from other pandemics

The other major pandemic during the past 20 years was the 2014 Ebola outbreak in West Africa. See Figure 4. Since Ebola is not airborne, its implications for COVID-19 are limited. However, it is worth noting some of the predictions made at the time and the lessons learned. A number of models addressed the Ebola epidemic in its earliest stages.

Rivers *et al.* [52] investigated the efficiency of increased contact tracing, improved infection-control practices and a hypothetical pharmaceutical intervention to improve survival of hospitalised patients aimed at reducing the reproduction number of the Ebola virus. They showed that perfect contact tracing could reduce R_0 from 2.22 to 1.89, while additionally reducing hospitalisation rates by 75% would further reduce R_0 to 1.72. Fisman *et al.* [53] used a two-parameter model to examine epidemic growth and control. They evaluated the growth patterns and determined the degree to which the epidemic could be controlled, finding only weak control in West Africa as of the end of August 2014 and essentially no control in Liberia. Browne *et al.* [54] developed an SEIR model to examine the effects of contact tracing. They showed that if contact tracing was perfect, then the critical proportion of contacts that need to be traced could be derived. Webb and Browne [55] developed an age-structured model, tracking disease age through initial incubation, followed by an infectious phase with variable transmission infectiousness. They showed that successive stages of hospitalisation would result in a mitigation of the epidemic. Do and Lee [56] developed an SLIRD model to determine the mathematical dynamics of the disease. Their model had a single equilibrium point, which they showed would be stable if safe burial practices without traditional rituals were followed and $R_0 < 1$. Salem and Smith? [57] conducted a sensitivity analysis, showing safe burial and reduced contacts — with both living and dead infected individuals — were the interventions most likely to lead to eradication. However, models for the 2014 Ebola epidemic tended to significantly overestimate the the number of cases and deaths [58].

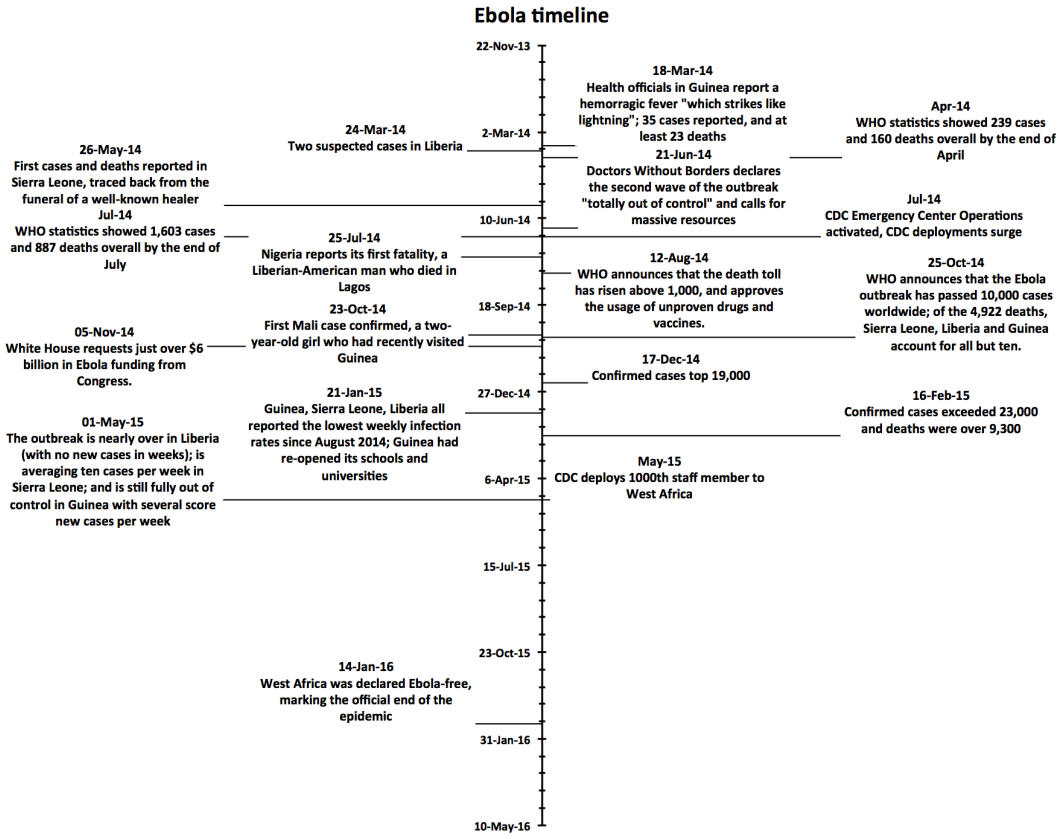

Figure 4. Ebola timeline

In addition to specific diseases, mathematical models have also made general predictions for pandemics not linked to any specific disease or classified under the umbrella of pandemic influenza. Such models are numerous and not always based on data or strict biological accuracy, but it is worth acknowledging that some have made insightful forecasts, nonetheless. In 2006, Germann *et al.* [59] used stochastic modelling to describe initial seeding and final quenching of small community-level outbreaks, with a suggested range of R_0 values from 1.6 to 2.4. Intervention methods considered were socially targeted antiviral prophylaxis, dynamic mass vaccination, closure of schools and social distancing. They considered mitigation strategies to be successful when the attack rate was limited to approximately 10% of the US population, in line with that of annual influenza epidemics. In 2011, Wu and Cowling [60] surveyed the state of mathematical models predicting influenza pandemics. They suggested that travel restrictions would have little effect in stopping the pandemic but that the most effective mitigation would be combinations of antiviral treatment and prophylaxis, as well as non-pharmaceutical interventions such as quarantine, isolation, school closure and social distancing in both the community and workplace. Short-term solutions would be maintaining a stockpile of antiviral drugs, while a long-term solution would be the development of a vaccine. In 2012, Morse *et al.* [61] examined the emergence of novel infectious diseases from zoonoses, which account for more than 60% of the 400 emerging infectious diseases that have been identified since 1940. They noted that no novel pathogens have ever been predicted before their first appearance but that predictions were possible once an outbreak was underway. They identified three stages of a new pandemic: 1) no hu-

man infection; 2) localised human infection; and 3) widespread transmission and global dissemination. In 2013, Huppert and Katriel [62] argued that mathematical models can and do provide accurate predictions in the field of infectious diseases, provided the scope of the notion of prediction is suitably qualified. They suggested that the next pandemic would cause severe morbidity, mortality and extensive economic impact and that the advantages of modern medicine would be offset by the ability of transportation networks to rapidly propagate a disease around the globe. In 2017, Saunders-Hastings *et al.* [63] examined strategies for mitigating a potential influenza pandemic and calculated economic costs, with hospital-resource capacity as a limiting factor. In addition to pharmaceutical interventions such as antivirals and vaccination, the non-pharaceutical interventions under consideration were school closures, community-contact reduction, hand hygeine, mask use, voluntary isolation and quarantine.

It is worth noting that all of the above non-pharmaceutical predictions came to pass during the COVID-19 outbreak, while the pharmaceutical-based predictions may yet come to do so if a vaccine is developed. This illustrates the power of modelling to make good generalised predictions even in the absence of specific data

6.2. Implications for COVID-19

The lessons learned from early modelling of SARS, MERS and the 2009 H1N1 outbreak show that it is possible to derive meaningful estimates for disease spread, the maximum number of infections and an approximate fade-away date when using appropriate model selection and characteristics. Stochastic models that take into account the variability of disease spread are extremely useful, both in the early asymptomatic stages and through its regular course. Where possible, spatial considerations seem to help the accuracy when outbreaks occur in highly populated, dense regions like Hong Kong. Allowing for the possibility of SSEs is paramount. To create accurate models, it is imperative to first determine the nature of disease spread between individuals that will make up the pathways between compartments in order to answer the following questions: Can individuals become reinfected? Can recovered individuals continue to spread the disease? As seen in the case of MERS, it is possible to closely estimate very short-term, powerful and volatile events of disease spread when the nature of the disease and setting (e.g., hospitals) are very similar to those of known pandemics. Significant difficulty might arise if the disease appears in large, consecutive waves.

COVID-19 has many similarities with the recent outbreaks, but there are important differences. Most obvious is the scale of the outbreak, which is significantly greater than anything witnessed in the past century. Whereas SARS was entirely contained to its first wave and MERS varied between low-level community transmission and rapid outbreaks in hospitals, both the 2009 H1N1 outbreak and COVID-19 produced multiple sustained waves. An effective vaccine was developed later in the 2009 H1N1 outbreak [64]; at the time of writing, it remains to be seen whether a vacine for COVID-19 will reach the market, and if so how effective it will be. Lessons from early Ebola modelling illustrated the importance of effective contact tracing and the reduction of contacts with infected individuals, which are also important factors for COVID-19.

It follows that current and future pandemics can be modelled with some degree of accuracy, at least in the first wave. Understanding how a pandemic might unfold during its earliest stages is clearly highly desirable, and the ability of early COVID-19 models to predict the disease's course is possible, as seen from the lessons learned from SARS, MERS and the 2009 H1N1 outbreak. It follows that models for COVID-19 and future pandemics are most useful when they include stochasticity, spatial

considerations and SSEs. Furthermore, waves after the first should be treated as distinct waves for the purpose of modelling. Mathematical modelling can be a powerful tool for predicting the future when harnessed correctly.

Acknowledgements

EJS is supported by the Simons Foundation. RS? is supported by an NSERC Discovery Grant and an NSERC Alliance Grant. The authors are grateful to two anonymous reviewers, whose comments helped improve the manuscript. For citation purposes, please note that the question mark in "Smith?" is part of his name.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Y. Liu, A. A. Gayle, A. Wilder-Smith and J. Rocklöv, The Reproductive Number of COVID-19 Is Higher Compared to SARS Coronavirus, *Journal of Travel Medicine* **27** (2020), taaa021.
2. Q. Li. X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong and X. Xing, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia, *New England Journal of Medicine*, **382** (2020), 1199–1207.
3. W. Wang, J. Tang and F. Wei, Updated understanding of the outbreak of 2019 novel coronavirus (2019nCoV) in Wuhan, China, *Journal of Medical Virology* **92** (2020), 441–447.
4. K. Kousha and M. Thelwall, COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts, *Quantitative Science Studies* **1** (2020), 1068–1091.
5. J. M. Heffernan, R. J. Smith and L. M. Wahl, Perspectives on the basic reproductive ratio, *Journal of the Royal Society Interface* **2** (2005), 281–293.
6. J. Li, D. Blakeley and R. J. Smith?, The Failure of R_0, *Comp Math Methods Med*, **2011** (2011), 527610.
7. B. Tang, X. Wang, Q. Li, N. L. Bragazzi, S. Tang, Y. Xiao and J. Wu, Estimation of the Transmission Risk of the 2019-NCoV and Its Implication for Public Health Interventions, *Journal of Clinical Medicine*, **9** (2020), 462.
8. N. Imai, A. Cori, I. Dorigatti, M. Baguelin, C. Donnelly, S. Riley and N. Ferguson, Report 3: Transmissibility of 2019-nCoV, *Imperial College London* (2020) 1–6.
9. S. Zhao, Q. Lin, J. Ran, S. S. Musa, G. Yang, W. Wang, Y. Lou, D. Gao, L. Yang, D. He and M. H. Wang, Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak, *International Journal of Infectious Diseases*, **92** (2020), 214–217.
10. J. T. Wu, K. Leung and G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *The Lancet*, **395** (2020), 689–697.

11. J. Riou and C. L. Althaus, Pattern of early human-to-human transmission of Wuhan 2019-nCoV, *Eurosurveillance*, **25** (2020), pii=2000058.
12. R. D. Smith, Responding to Global Infectious Disease Outbreaks: Lessons from SARS on the Role of Risk Perception, Communication and Management. *Social Science and Medicine*, **63** (2006), 3113–3123.
13. World Health Organization, Summary of Probable SARS Cases with Onset of Illness from 1 November 2002 to 31 July 2003, https://www.who.int/csr/sars/country/table2004_04_21/en/. Accessed 13 Oct 2020.
14. S. Riley, C. Fraser, C. A. Donnelly, A. C. Ghani, L. J. Abu-Raddad, A. J. Hedley, G. M. Leung, L. M. Ho, T. H. Lam, T. Q. Thach and P. Chau, Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions, *Science*, **300** (2003), 1961–1966.
15. G. Chowell, P. W. Fenimore M. A. Castillo-Garsow and C. C. Castillo-Chavez, SARS Outbreaks in Ontario, Hong Kong and Singapore: The Role of Diagnosis and Isolation as a Control Mechanism, *Journal of Theoretical Biology*, **224** (2003), 1–8.
16. M. Lipsitch, T. Cohen, B. Cooper, J. M. Robins, S. Ma, L. James, G. Gopalakrishna, S. K. Chew, C. C. Tan, M. H. Samore and D. Fisman, Transmission Dynamics and Control of Severe Acute Respiratory Syndrome, *Science*, **300** (2003), 1966–1970.
17. G. Zhou and G. Yan. Severe Acute Respiratory Syndrome Epidemic in Asia, *Emerging Infectious Diseases*, **9** (2003), 1608–1610.
18. B. C. K. Choi and A. W. P. Pak. A Simple Approximate Mathematical Model to Predict the Number of Severe Acute Respiratory Syndrome Cases and Deaths, *Journal of Epidemiology and Community Health*, **57** (2003), 831–835.
19. L. O. Lloyd-Smith, A. P. Galvani, W. M. Getz, Curtailing Transmission of Severe Acute Respiratory Syndrome within a Community and Its Hospital, *Proceedings of the Royal Society of London. Series B: Biological Sciences*, **270**, (2003), 1979–1989.
20. S. A. Eifan, I. Nour, A. Hanif, A. M. Zamzam, S. M. AlJohani, A Pandemic Risk Assessment of Middle East Respiratory Syndrome Coronavirus (MERS-CoV) in Saudi Arabia, *Saudi Journal of Biological Sciences*, **24** (2017), 1631–1638.
21. G. Chowell, F. Abdirizak, S. Lee, J. Lee, E. Jung, H. Nishiura and C. Viboud. Transmission Characteristics of MERS and SARS in the Healthcare Setting: A Comparative Study, *BMC Medicine***13** (2015), 210.
22. C. Drosten, B. Meyer, M. A. Müller, V. M. Corman, M. Al-Masri, R. Hossain, H. Madani, A. Sieberg, B. J. Bosch, E. Lattwein and R. F. Alhakeem, Transmission of MERS-coronavirus in household contacts, *New England Journal of Medicine*, **371** (2014), 828–835.
23. S. Cauchemez, C. Fraser, M. D. Van Kerkhove, C. A. Donnelly, S. Riley, A. Rambaut, V. Enouf, S. van der Werf and N. M. Ferguson, Middle East Respiratory Syndrome Coronavirus: Quantification of the Extent of the Epidemic, Surveillance Biases, and Transmissibility, *The Lancet Infectious Diseases*, **14** (2014), 50–56.
24. R. Breban, J. Riou, A. Fontanet, Interhuman Transmissibility of Middle East Respiratory Syndrome Coronavirus: Estimation of Pandemic Risk, *The Lancet*, **382** (2013), 694–699.

25. H.-J. Chang, Estimation of Basic Reproduction Number of the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) during the Outbreak in South Korea, 2015, *BioMedical Engineering OnLine*, **16** (2017), 79.
26. Centers for Disease Control, H1N1 Flu Pandemic Timeline. Centers for Disease Control and Prevention, https://www.cdc.gov/flu/pandemic-resources/2009-pandemic-timeline.html Accessed 13 Oct 2020.
27. T. N. Jilani, R. T. Jamil and A. H. Siddiqui, H1N1 Influenza (Swine Flu). *StatPearls*, 2020, http://www.ncbi.nlm.nih.gov/books/NBK513241/ Accessed 13 Oct 2020.
28. Centers for Disease Control, 2009 H1N1 Pandemic (H1N1pdm09 virus), https://www.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html Accessed 13 Oct 2020.
29. M. G. Roberts and H. Nishiura. Early Estimation of the Reproduction Number in the Presence of Imported Cases: Pandemic Influenza H1N1-2009 in New Zealand *PLoS ONE*, **6** (2011), e17835.
30. C. Fraser, C. A. Donnelly, S. Cauchemez, W. P. Hanage, M. D. Van Kerkhove, T. D. Hollingsworth, J. Griffin, R. F. Baggaley, H. E. Jenkins, E. J. Lyons and T. Jombart, Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings, *Science*, **324** (2009), 1557–1561.
31. H. Nishiura, C. Castillo-Chavez, M. Safan M and G. Chowell, Transmission Potential of the New Influenza A(H1N1) Virus and Its Age-Specificity in Japan, *Eurosurveillance*, **14** (2009), 19227.
32. L. F. White, J. Wallinga, L. Finelli, C. Reed, S. Riley, M. Lipsitch and M. Pagano, Estimation of the Reproductive Number and the Serial Interval in Early Phase of the 2009 Influenza A/H1N1 Pandemic in the USA. *Influenza and Other Respiratory Viruses*, **3** (2009), 267–276.
33. L. C. Mostaço-Guidolin, C. S. Bowman, A. L. Greer, D. N. Fisman and S. M. Moghadas, Transmissibility of the 2009 H1N1 Pandemic in Remote and Isolated Canadian Communities: A Modelling Study. *BMJ Open*, **2** (2012), e001614.
34. M. Helferty, J. Vachon, J. Tarasuk, R. Rodin, J. Spika and L. Pelletier, Incidence of Hospital Admissions and Severe Outcomes during the First and Second Waves of Pandemic (H1N1) 2009. *CMAJ*, **182** (2010), 1981–1987.
35. O. T. Mytton, P. D. Rutter, M. Mak, E. A. Stanton, N. Sachedina and L. J. Donaldson, Mortality Due to Pandemic (H1N1) 2009 Influenza in England: A Comparison of the First and Second Waves, *Epidemiology and Infection*, **140** (2012), 1533–1541.
36. I. Dorigatti, S. Cauchemez, N. M. Ferguson, Increased Transmissibility Explains the Third Wave of Infection by the 2009 H1N1 Pandemic Virus in England, *Proceedings of the National Academy of Sciences of the United States of America*, **110** (2013), 13422–13427.
37. O. Sharomi, C. N. Podder, A. B. Gumel, S. M. Mahmud and E. Rubinstein, Modelling the Transmission Dynamics and Control of the Novel 2009 Swine Influenza (H1N1) Pandemic, *Bulletin of Mathematical Biology*, **73** (2011), 515–548.
38. M. A. Jhung, D. Swerdlow, S. J. Olsen, D. Jernigan, M. Biggerstaff, L. Kamimoto, K. Kniss, C. Reed, A. Fry, L. Brammer, J. Gindler, W. J. Gregg, J. Bresee and L. Finelli L, Epidemiology of 2009 pandemic influenza A (H1N1) in the United States, *Clinical Infectious Diseases*, **52** (2011), S13–S26.

39. M.D. Van Kerkhove, A. W. Mounts, S. Mall, K. A. Vandemaele, M. Chamberland, T. dos Santos, J. Fitzner, M. A. Widdowson, J. Michalove, J. Bresee and S. J. Olsen, Epidemiologic and virologic assessment of the 2009 influenza A (H1N1) pandemic on selected temperate countries in the Southern Hemisphere: Argentina, Australia, Chile, New Zealand and South Africa, *Influenza Other Respiratory Viruses*, **5** (2011), e487–e498.
40. R. J. Smith?, Did we Eradicate SARS? Lessons Learned and the Way Forward, *American Journal of Biomedical Science & Research*, **6** (2019), 001017.
41. R. M. Anderson, C. Fraser, A. C. Ghani, C. A. Donnelly, S. Riley, N. M. Ferguson, G. M. Leung, T. H. Lam and A. J. Hedley, Epidemiology, Transmission Dynamics and Control of SARS: The 2002–2003 Epidemic, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **359** (2004), 1091–105.
42. S. Ruan, W. Wang and S. A. Levin, The Effect of Global Travel on the Spread of SARS, *Mathematical Biosciences and Engineering*, **3** (2006), 205–218.
43. N. G. Becker, K. Glass, Z. Li and G. K. Aldis, Controlling Emerging Infectious Diseases like SARS, *Mathematical Biosciences*, **193** (2005), 205–121.
44. A. J. Kucharski and C. L. Althaus, The role of superspreading in Middle East respiratory syndrome coronavirus (MERS-CoV) transmission *Eurosurveillence* **20** (2015), 14–18.
45. S. Bernard-Stoecklin, B. Nikolay, A. Assiri, A. A. Saeed, P. K. Embarek, H. El Bushra, M. Ki, M. R. Malik, A. Fontanet, S. Cauchemez and M. D. Van Kerkhove, Comparative Analysis of Eleven Healthcare-Associated Outbreaks of Middle East Respiratory Syndrome Coronavirus (Mers-Cov) from 2015 to 2017, *Scientific Reports*, **9** (2019), 1–9.
46. S. Choi, E. Jung, B. Y. Choi, Y. J. Hur and M. Ki, High Reproduction Number of Middle East Respiratory Syndrome Coronavirus in Nosocomial Outbreaks: Mathematical Modelling in Saudi Arabia and South Korea, *Journal of Hospital Infection*, **99** (2018), 162–168.
47. T. Sardar, I. Ghosh, X. Rodó and J. Chattopadhyay, A Realistic Two-Strain Model for MERS-CoV Infection Uncovers the High Risk for Epidemic Propagation, *PLOS Neglected Tropical Diseases*, **14** (2020), e0008065.
48. M.S. Majumder, C. Rivers, E. Lofgren and D. Fisman. Estimation of MERS-Coronavirus Reproductive Number and Case Fatality Rate for the Spring 2014 Saudi Arabia Outbreak: Insights from Publicly Available Data, *PLoS Currents*, **6** (2014).
49. P. Poletti, M. Ajelli and S. Merler, The effect of risk perception on the 2009 H1N1 pandemic influenza dynamics, *PLoS One*, **6** (2011), e16460.
50. S. Tsukui, Case-Based Surveillance of Pandemic (H1N1) 2009 in Maebashi City, Japan, *Japanese Journal of Infectious Diseases*, **65** (2012), 132–137.
51. G. Chowell, S. Echevarria-Zuno, C. Viboud, L. Simonsen, J. Tamerius, M. A. Miller and V. H. Borja-Aburto, Characterizing the epidemiology of the 2009 influenza A/H1N1 pandemic in Mexico, *PLoS Medicine*, **8** (2011), e1000436.
52. C.M. Rivers, E. T. Lofgren, M. Marathe, S. Eubank, and B. L. Lewis, Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia, *PLOS Currents Outbreaks*, **6** (2014).

53. D. Fisman, E. Khoo and A. Tuite, Early epidemic dynamics of the West African 2014 Ebola outbreak: Estimates derived with a simple two-parameter model. *PLoS Currents*, **6** (2014).

54. C. Browne, H. Gulbudak and G. Webb, Modeling contact tracing in outbreaks with application to Ebola. *Journal of Theoretical Biology*, **384** (2015), 33–49.

55. G. Webb and C. Browne, A model of the Ebola epidemics in West Africa incorporating age of infection. *Journal of Biological Dynamics*, **10** (2016), 18–30.

56. T. S. Do and Y. S. Lee, Modeling the spread of Ebola. *Osong Public Health and Research Perspectives*, **7** (2016), 43–48.

57. D. Salem and R. Smith?, A Mathematical Model of Ebola Virus Disease: Using Sensitivity Analysis to Determine Effective Intervention Targets, *Proceedings of the SummerSim-SCSC 2016 conference*, (2016), 16–23 .

58. P. Bhandari, Analysis of Prediction Models in spread of Ebola Virus Disease, *Thesis, Deakin University* (2019).

59. T. C. Germann, K. Kadau, I. M. Longini, C. A. Macken, Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences*, **103** (2006), 5935–5940.

60. J. T. Wu and B. J. Cowling, The use of mathematical models to inform influenza pandemic preparedness and response, *Experimental Biology and Medicine*, **236** (2011), 955–961.

61. S. S. Morse, J. A. Mazet, M. Woolhouse, C. R. Parrish, D. Carroll, W. B. Karesh, C. Zambrana-Torrelio, W. I. Lipkin and P. Daszak, Prediction and prevention of the next pandemic zoonosis, *The Lancet*, **380** (2012),1956–1965.

62. A. Huppert and G. Katriel, Mathematical modelling and prediction in infectious disease epidemiology. *Clinical microbiology and infection*, **19** (2013), 999–1005.

63. P. Saunders-Hastings, B. Q. Hayes, R. Smith? and D. Krewski. Modelling community-control strategies to protect hospital resources during an influenza pandemic in Ottawa, Canada. *PloS One*, **12** (2017), e0179315.

64. M. Valenciano, E. Kissling, J. M. Cohen, N. Oroszi, A. S. Barret, C. Rizzo, B. Nunes, D. Pitigoi, A. L. Cámara, A. Mosnier and J. K. Horvath. Estimates of pandemic influenza vaccine effectiveness in Europe, 2009–2010: results of Influenza Monitoring Vaccine Effectiveness in Europe (I-MOVE) multicentre case-control study. *PLoS Medicine*, **8** (2011), e1000388.

AIMS Press