

Modelling in infectious diseases: between haphazard and hazard

A. Neuberger^{1,2}, M. Paul^{1,3}, A. Nizar² and D. Raoult⁴

1) Unit of Infectious Diseases, Rambam Health Care Campus, Haifa, 2) Department of Medicine B, Rambam Health Care Campus, Haifa, 3) Sackler Faculty of Medicine, Tel-Aviv University, Ramat-Aviv, Israel and 4) Unité de Recherche en Maladies Infectieuses et Tropicales Émergentes, CNRS-IRD UMR 6236, Faculté de Médecine, Université de la Méditerranée, Marseille, France

Abstract

Modelling of infectious diseases is difficult, if not impossible. No epidemic has ever been truly predicted, rather than being merely noticed when it was already ongoing. Modelling the future course of an epidemic is similarly tenuous, as exemplified by ominous predictions during the last influenza pandemic leading to exaggerated national responses. The continuous evolution of microorganisms, the introduction of new pathogens into the human population and the interactions of a specific pathogen with the environment, vectors, intermediate hosts, reservoir animals and other microorganisms are far too complex to be predictable. Our environment is changing at an unprecedented rate, and human-related factors, which are essential components of any epidemic prediction model, are difficult to foresee in our increasingly dynamic societies. Any epidemiological model is, by definition, an abstraction of the real world, and fundamental assumptions and simplifications are therefore required. Indicator-based surveillance methods and, more recently, Internet biosurveillance systems can detect and monitor outbreaks of infections more rapidly and accurately than ever before. As the interactions between microorganisms, humans and the environment are too numerous and unexpected to be accurately represented in a mathematical model, we argue that prediction and model-based management of epidemics in their early phase are quite unlikely to become the norm.

Keywords: Epidemics, epidemiology, models, pandemic, prediction

Article published online: 25 June 2013

Clin Microbiol Infect 2013; **19**: 993–998

Corresponding author: A. Neuberger, Unit of Infectious Diseases and Internal Medicine B, Rambam Medical Centre, Haalyah Hashnyah 8 St, Haifa, Israel

E-mail: a_neuberger@rambam.health.gov.il

Introduction

Prophecy is a good line of business, but it is full of risks.

Mark Twain in *Following the Equator*

Epidemics have played a role in human history since ancient times, and will continue to do so in the foreseeable future, despite overoptimistic assurances to the contrary. When the Black Death pandemic was ravaging Europe during the Middle Ages, the only sound advice given to the citizens was “flee early, flee far, return late”. As reflected in the Introduction of

Boccaccio's *Decameron*, the citizens of Florence “decided that the only remedy for the pestilence was to avoid it ... [that] none ought to stay in a place thus doomed to destruction”. Modern medicine does not have to resort to such extreme measures of public health, and nor does it ascribe the occurrence of epidemics to a certain alignment of the stars, the will of God, harmful vapours, or the poisoning of wells by non-believers. Surveillance systems make the early detection of disease outbreaks possible through data supplied by sentinel clinics, or by the use of syndromic surveillance (e.g. Web queries, other forms of Internet biosurveillance, over-the-counter drug sales, or school absence records) [1]. We would

argue, however, that timely prediction of epidemics before they occur, and accurate forecasts of their course in their early phase, remain, by and large, unreliable.

Pathogen–pathogen Interactions and other Unknowns

The Division of Tuberculosis Control shares the belief of the symposium participants that tuberculosis will virtually disappear in the United States in the next 50 years. The control and eradication of tuberculosis, *New England Journal of Medicine*, 1980.

Since the early 1950s, tuberculosis (TB) rates in high-income countries have decreased rapidly. In 1980, treatment was available and effective, and it seemed reasonable to include TB in the list of “disappearing and declining diseases” in Britain [2]. The authors were naturally unaware of the fact that the AIDS pandemic was already making hundreds of thousands of people worldwide susceptible to a disease previously considered to be a remnant of the 19th century [3]. The surge in the incidence of TB and the appearance of multidrug-resistant and extensively drug-resistant TB in eastern Europe in recent decades is causally linked to a wide variety of actors: the AIDS pandemic, the collapse of the Soviet Union, and the increase in intravenous drug abuse there. Predicting the occurrence of these epidemiological and political phenomena was not possible in 1980. Beijing genotype strains of *Mycobacterium tuberculosis* now account for approximately 50% of TB cases in China, and are spreading worldwide [4]. This genotype has been observed to spread more successfully in the population than other *M. tuberculosis* strains. The reasons for this are incompletely understood [4]. Will this genotype change the epidemiology of TB? Will vaccination and treatment trigger the appearance of other successful genotypes? Nearly 20 years after the first description of the *M. tuberculosis* Beijing genotype, and more than 30 years after the recurrence of the TB pandemic, we still lack elementary biological and epidemiological data to help with TB control.

The association between influenza virus infection and subsequent susceptibility to *Streptococcus pneumoniae* infections was already well known nearly 100 years ago during the Spanish influenza pandemic. However, the virus itself has the capacity to mutate, and human society changes continuously, so that predictions cannot be based on such historical associations. If predictions were to rely on past observations, one would expect adults aged ≥ 65 years, who are known to be susceptible to both severe influenza and pneumococcal infections, to have extremely high mortality rates during influenza pandemics. Hygienic conditions today, however, are

different from those of 1919, pneumococcal sepsis being the exception among patients with influenza. During the 2009 A/H1N1 influenza virus pandemic, patients aged ≥ 65 years were found to have death rates 81% lower than expected in a regular influenza season [5]. In fact, obesity, not ageing, was found to be a significant risk factor for severe disease [5,6]. Not only was the influenza virus itself different, but it also interacted with other viruses in important ways. In France, for instance, a rhinovirus epidemic was found to delay the onset of the influenza pandemic, which in itself delayed the onset of the respiratory syncytial virus bronchiolitis season [6–8]. To complicate things further, increasing evidence suggests that some bacterial infections can also increase the susceptibility of patients to viral infections [9].

Pathogen–environment Interactions

Not only the pathogens themselves, but also the complex ecosystems, which include vectors and/or reservoir animals, are crucial to understanding the dynamics of many infectious diseases with an epidemic potential. The interaction between the *Anopheles* mosquito vector, the *Plasmodium falciparum* parasite and humans is a good example of such complexity. In recent years, long-lasting insecticide-treated bed-nets have been distributed in many sub-Saharan African countries, following evidence from randomized controlled trials that these reduce *P. falciparum* malaria prevalence, morbidity, and mortality [10]. Although it was reasonable to assume that the continuation of such efforts would lead to a gradual and predictable decrease in malaria morbidity, the results of a recent longitudinal study performed in Senegal highlight the problematic nature of such simplistic forecasts. In this study, the average incidence density of malaria attacks, which was 5.45 per 100 person-months before the distribution of treated bed-nets, decreased to 0.41 immediately afterwards, only to increase again to 4.57 per 100 person-months 27–20 months after the initial intervention, despite continued use of the bed-nets. The prevalence of knockdown resistance mutation, which confers reduced sensitivity of the *Anopheles* vector to pyrethroid insecticides, increased from 8% in 2007 to 48% in 2010. The mosquitoes were shown to become somewhat more aggressive during the early evening, thereby avoiding the need to ‘confront’ bed-nets [11]. Unpredictable events such as these undermine the various attempts to model and predict trends in malaria control and eradication [12,13].

There has been no cholera epidemic in the Caribbean island of Hispaniola for more than a century, although cholera has been present in Latin America since 1991. The *Vibrio cholerae* strain that spread to all Haitian provinces after the 2010

earthquake originated in Asia, and not from the neighbouring countries in the Americas [14]. It has been suggested that the bacteria were introduced into Haiti by United Nations soldiers sent to Haiti after the earthquake. If this was indeed the case, prediction of a cholera epidemic in Haiti in 2010 would also have required, in addition to all other factors, an accurate earthquake forecast.

The Ever-changing Variables

Contrariwise, continued Tweedledee, if it was so, it might be;
and if it were so, it would be; but as it isn't, it ain't. That's logic.

Lewis Carroll in *Through the Looking Glass*

The list of factors that need to be included in an 'ideal' model of epidemic prediction seems never-ending. We choose to include certain variables in a model, but deliberately or inadvertently ignore others.

Human-related variables include population density, nutritional status, the number of susceptible hosts within the population, infection control measures taken by individuals within the population, healthcare infrastructure and available resources, domestic and international travel, the use and impact of quarantine, the use of (or refusal to use) antimicrobials and vaccines, and the public reaction to the epidemic (e.g. population migration and closure of schools). Human African trypanosomiasis (HAT), for example, was considered to be a candidate for eradication in most African countries in the 1960s. The disease re-emerged later on, despite the availability of effective, if somewhat toxic, treatment options. The reasons for the increase in HAT incidence included factors such as political instability in some African countries, such as the Democratic Republic of Congo and the Central African Republic, failing healthcare systems, neglect of existing HAT diagnosis and vector control programmes, lack of investment in new drug development by pharmaceutical companies, and the consideration of withdrawal of existing drugs on economic grounds [15]. It is doubtful whether any of the above factors could be reliably represented in a mathematical model.

Pathogen-related variables include, but are not limited to, the duration of an incubation period, the period of pathogen infectivity, the rate of disease transmission, the average age at which a disease is typically contracted in a given population, virulence, the susceptibility of the organism to antimicrobials, and the availability of a vaccine. Most emerging infectious diseases and nearly all pandemics were caused by 'classic' zoonotic pathogens (e.g. the plague), or by pathogens that were initially confined to animals or had limited potential for

causing human infections, but then mutated, crossed the species barrier, and disseminated globally (e.g. SARS coronavirus, and human immunodeficiency virus) [16]. A list containing animal pathogens that could potentially trigger a new pandemic can be compiled; surveillance aimed at detecting an outbreak of a disease caused by one of these pathogens should hence be continuous. However, although making predictions even about an organism as extensively researched as the influenza virus seems to be difficult, accurately predicting the course of a pandemic caused by a pathogen newly introduced into a human population is, in all likelihood, nearly impossible.

Finally, climate and environmental changes and their effects on humans, intermediate hosts, reservoir animals and vectors are all instrumental in the understanding of many infectious diseases [17]. They are, however, still poorly understood, immensely complex, and difficult to predict with any accuracy—as are, for instance, the changing seasonality of influenza epidemics, the accelerated transmission of certain West Nile virus genotypes with increasing temperatures, and the predicted extension northwards of freshwater snail-mediated schistosomiasis in China [6,18,19]. Until several years ago, malaria was considered to be in the pre-elimination phase in Malaysia. The zoonotic *Plasmodium knowlesi*, a parasite that mainly infects monkeys, has emerged as the dominant malaria species in Malaysian Borneo, and is increasingly being reported in other countries. Despite the fact that deforestation, increased human activity at the forest fringes, rapid growth in the population of Malaysian Borneo and closer contact between humans and macaques were quite predictable, the emergence of *P. knowlesi* as major human malaria parasite was noted only in retrospect [20].

All of these factors are not only constantly changing, but are also interacting with each other in an infinite number of ways. Thus, it is no wonder that no new epidemic has ever been truly predicted, rather than being merely noticed when it was already ongoing.

Models

As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality.

Albert Einstein

The construction of a model aimed at predicting the course of an epidemic necessitates assumptions and simplifications [21]. For example, it is common to assume a rectangular age distribution, with most individuals in a population reaching old age, a typical observation in high-income countries with low

infant mortality rates. The poor and the displaced, however, are the populations most affected by epidemics such as AIDS, TB, epidemic typhus, or cholera, and the age distribution of these populations is vastly different. Another simplification often used in the construction of deterministic models rests on the assumption of homogeneous mixing of the population, i.e. on the premise that all individuals in a certain population associate randomly with each other. A recent mumps epidemic in New York and New Jersey affected 3502 patients, 89% of whom were fully immunized. The assumption of homogeneous mixing of the population seems absurd, as 97% of cases occurred in the orthodox Jewish population group; 78% of them were male, and adolescents attending religious schools were disproportionately affected [22]. More complex stochastic models, which take variability and chance into account, contain less epidemiologically improbable assumptions, but are best applied to smaller populations, and require the inclusion of more variables, greatly complicating their use.

Many policy-makers, journalists and doctors lack an in-depth understanding of a model's structure or limitations, and may choose to accept or reject a certain prediction on the basis of non-rational causality. The 2003 SARS epidemic provides a good example. Early models suggested that isolation of contacts before symptom onset would be beneficial in controlling the spread of the disease [23]. Contact tracing and quarantine of asymptomatic people is, however, a daunting task for any public health system, and only approximately 5% of contacts who eventually became ill were in fact isolated before becoming symptomatic [24]. An analysis of the real-life impact of contact tracing and quarantine during the SARS crisis will, in all likelihood, show that timely isolation of symptomatic patients would have achieved nearly identical results, with much greater efficiency. Quarantine of asymptomatic contacts contributed little to SARS control, but probably led to excess costs, increased psychological stress among those quarantined, and a lingering misunderstanding of how the epidemic was actually contained [24].

During the SARS epidemic, several estimates of the basic reproduction number (i.e. R_0 , which is defined as the average number of people infected by one patient during an epidemic) were published, and were generally in the range of 2–5 [24,25]. If an R_0 of 2.6–3.2 had been used, one would have expected 30 000–10 000 000 SARS cases in China alone. Eventually, only 782 cases were reported, suggesting a much lower R_0 [26]. At the beginning of the SARS epidemic, during the course of a 'super-spreading event', one person infected as many as 300 people [25]. Eventually, the reproduction number dropped dramatically. Such a wide variation in R_0 values demonstrates how prone models are to errors when they use limited data available during the initial phases of an outbreak,

based mostly on case reports. Why did SARS disappear? Was it a huge success of international health regulations or a poorly understood phenomenon? Retrospectively, we lack the understanding to model this epidemic's course.

Doomsday predictions are more frequently discussed in the popular media, and are probably also more likely to be accepted for publication in scientific journals. The inadvertent promotion of fear is likely to attract public funds, and will, in retrospect, be applauded for any correct predictions, but forgotten when found to be incorrect. In 1966, the economist Paul A. Samuelson famously noted that "Wall Street indexes predicted nine out of the last five recessions". This observation is also relevant to yearly threats of a new pandemic, only a minority of which actually materialize. In 2009, the French emergency plan during the 2009 influenza pandemic was based on an estimation of 91 000–210 000 deaths, and led to the opening of 700 new hospital beds exclusively for influenza patients in Marseille. In reality, <300 patients were hospitalized, and no more than 50 beds were used at the same time, even at the pandemic's peak [6].

When epidemiological predictions are made, the data used are based on past observations, which may be irrelevant or inaccurate. In 1990, the *Journal of the American Medical Association* published an article describing the projected size of the AIDS epidemic based on Farr's law, which states that the rise and fall of an epidemic curve is roughly symmetrical and can be approximated by a normal bell-shaped curve [27]. William Farr, a British doctor and epidemiologist, based this model on his observations of smallpox and cholera epidemics in 19th-century London. The use of the same assumptions for AIDS, a disease that is different in nearly every epidemiological aspect, has led the authors to grossly underestimate AIDS incidence in the USA [28]. Several epidemiologists have noticed the flawed use of Farr's law, and a comment entitled 'AIDS Projections: How Farr Out?' was published in the same journal soon after [29]. The flawed prediction was, however, repeatedly cited by other authors.

Models' predictions usually have wide CIs, too wide to direct public health interventions. The estimated reproduction number (defined as the average number of secondary cases generated by a single infectious person) of the smallpox virus, a potential bioterrorist weapon, was used in constructing a model aimed at calculating the cumulative total number of smallpox cases after deliberate exposure [30]. The authors assumed that the transmission rate would be either 1.5 or 3.0 per person, that there would be an unlimited 'supply' of smallpox-susceptible persons, that exactly ten persons would initially be infected, and that no preventive intervention would be implemented. When transmission rates of either 1.5 or 3.0 were used, the number of individuals presumed to

become infected within 180 days fluctuated between 2190 and 2.2 million. After 365 days, the predicted number of infections was 224 000 vs. a theoretical 774 billion. These gargantuan differences illustrate the inability of models to provide accurate estimates *a priori*, even if several simplifying assumptions are made. Active surveillance of smallpox cases and the use of real-time data for model calibration will provide more reliable estimates. In the meantime, models describing bioterrorist smallpox attacks will yield doomsday scenarios for pessimists, and 'merely' unpleasant public health nuisance scenarios for optimists.

In conclusion, we argue that the interactions between microorganisms and humans are far too complex to be predictable. Most models used in epidemiological research still concentrate on one, known, pathogen, which causes a single disease in a well-defined population. It seems, however, that the reality at the microorganism level is much more complex, as organisms not only mutate, but continuously interact with the environment and with a large number of other organisms [31,32]. Indicator-based surveillance methods and, more recently, Internet biosurveillance systems can detect an outbreak of an infection more rapidly than ever before. Mathematical models play an important role in helping healthcare systems to respond to ongoing epidemics or plan the logistics of various theoretical scenarios, and were used in Haiti during the cholera epidemic, with real-time surveillance data [33]. Accurate predictions of epidemics before they occur are, however, quite unlikely to become the norm. Forecasting the course of epidemics is, to put it in Mark Twain's words, indeed "full of risks".

Acknowledgements

The authors thank A. Neuberger for editing the manuscript.

Transparency Declaration

All authors declare no conflicts of interest.

References

- Hulth A, Rydevik G, Linde A. Web queries as a source for syndromic surveillance. *PLoS One* 2009; 4: e4378.
- Galbraith NS, Forbes P, Mayon-White RT. Changing patterns of communicable disease in England and Wales: part II—Disappearing and declining diseases. *Br Med J* 1980; 281: 489–492.
- Gottlieb MS, Schroff R, Shanker HM et al. *Pneumocystis carinii* pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med* 1981; 305: 1425–1431.
- Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. *Lancet Infect Dis* 2010; 10: 103–111.
- Shrestha SS, Swerdlow DL, Borse RH et al. Estimating the burden of 2009 pandemic influenza A (H1N1) in the United States (April 2009–April 2010). *Clin Infect Dis* 2011; 52(suppl 1): S75–S82.
- Nougairède A, Charrel RN, Raoult D. Models cannot predict future outbreaks: A/H1N1 virus, the paradigm. *Eur J Epidemiol* 2011; 26: 183–186.
- Casalegno JS, Ottmann M, Duchamp MB et al. Rhinoviruses delayed the circulation of the pandemic influenza A (H1N1) 2009 virus in France. *Clin Microbiol Infect* 2010; 16: 326–329.
- Raoult D. Microbe interactions undermine predictions. *Science* 2011; 331: 144–145; author reply 5–7.
- Bosch AA, Biesbroek G, Trzcinski K, Sanders EA, Bogaert D. Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog* 2013; 9: e1003057.
- Lengeler C. Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database Syst Rev* 2004; 2: CD000363.
- Trape JF, Tall A, Diagne N et al. Malaria morbidity and pyrethroid resistance after the introduction of insecticide-treated bednets and artemisinin-based combination therapies: a longitudinal study. *Lancet Infect Dis* 2011; 11: 925–932.
- Reiner RC Jr, Perkins TA, Barker CM et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. *J R Soc Interface* 2013; 10: 20120921.
- Paul M. Are we losing the fight against malaria one more time? *Clin Microbiol Infect* 2011; 17: 1593–1596.
- Hendriksen RS, Price LB, Schupp JM et al. Population genetics of *Vibrio cholerae* from Nepal in 2010: evidence on the origin of the Haitian outbreak. *MBio* 2011; 2: e00157–11.
- Stich A, Barrett MP, Krishna S. Waking up to sleeping sickness. *Trends Parasitol* 2003; 19: 195–197.
- Morse SS, Mazet JA, Woolhouse M et al. Prediction and prevention of the next pandemic zoonosis. *Lancet* 2012; 380: 1956–1965.
- McMichael AJ. Globalization, climate change, and human health. *N Engl J Med* 2013; 368: 1335–1343.
- Kilpatrick AM, Meola MA, Moudy RM, Kramer LD. Temperature, viral genetics, and the transmission of West Nile virus by *Culex pipiens* mosquitoes. *PLoS Pathog* 2008; 4: e1000092.
- Zhou XN, Yang GJ, Yang K et al. Potential impact of climate change on schistosomiasis transmission in China. *Am J Trop Med Hyg* 2008; 78: 188–194.
- Barber BE, William T, Dhararaj P et al. Epidemiology of *Plasmodium knowlesi* malaria in north-east Sabah, Malaysia: family clusters and wide age distribution. *Malar J* 2012; 11: 401.
- Wearing HJ, Rohani P, Keeling MJ. Appropriate models for the management of infectious diseases. *PLoS Med* 2005; 2: e174.
- Barskey AE, Schulte C, Rosen JB et al. Mumps outbreak in Orthodox Jewish communities in the United States. *N Engl J Med* 2012; 367: 1704–1713.
- Lipsitch M, Cohen T, Cooper B et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science* 2003; 300: 1966–1970.
- Glasser JW, Hupert N, McCauley MM, Hatchett R. Modeling and public health emergency responses: lessons from SARS. *Epidemics* 2011; 3: 32–37.
- Dye C, Gay N. Epidemiology. Modeling the SARS epidemic. *Science* 2003; 300: 1884–1885.
- Louz D, Bergmans HE, Loos BP, Hoeben RC. Emergence of viral diseases: mathematical modeling as a tool for infection control, policy and decision making. *Crit Rev Microbiol* 2010; 36: 195–211.

27. Bregman DJ, Langmuir AD. Farr's law applied to AIDS projections. *JAMA* 1990; 263: 1522–1525.
28. Nishiura H. Lessons from previous predictions of HIV/AIDS in the United States and Japan: epidemiologic models and policy formulation. *Epidemiol Perspect Innov* 2007; 4: 3. DOI: 10.1186/1742-5573-4-3.
29. Artzrouni M. AIDS projections: how Farr out? *JAMA* 1990; 264: 1103–1104.
30. Meltzer MI, Damon I, LeDuc JW, Millar JD. Modeling potential responses to smallpox as a bioterrorist weapon. *Emerg Infect Dis* 2001; 7: 959–969.
31. Telfer S, Lambin X, Birtles R *et al.* Species interactions in a parasite community drive infection risk in a wildlife population. *Science* 2010; 330: 243–246.
32. Raoult D. Molecular, epidemiological, and clinical complexities of predicting patterns of infectious diseases. *Front Microbiol* 2011; 2: 25.
33. Abrams JY, Copeland JR, Tauxe RV *et al.* Real-time modelling used for outbreak management during a cholera epidemic, Haiti, 2010–2011. *Epidemiol Infect* 2013; 141: 1276–1285.