

Modelling microbial infection to address global health challenges

Meagan C. Fitzpatrick^{1,2}, Chris T. Bauch³, Jeffrey P. Townsend^{4,5,6} and Alison P. Galvani^{2,5,6*}

The continued growth of the world's population and increased interconnectivity heighten the risk that infectious diseases pose for human health worldwide. Epidemiological modelling is a tool that can be used to mitigate this risk by predicting disease spread or quantifying the impact of different intervention strategies on disease transmission dynamics. We illustrate how four decades of methodological advances and improved data quality have facilitated the contribution of modelling to address global health challenges, exemplified by models for the HIV crisis, emerging pathogens and pandemic preparedness. Throughout, we discuss the importance of designing a model that is appropriate to the research question and the available data. We highlight pitfalls that can arise in model development, validation and interpretation. Close collaboration between empiricists and modellers continues to improve the accuracy of predictions and the optimization of models for public health decision-making.

Microbial pathogens are responsible for more than 400 million years of life lost annually across the globe, a higher burden than either cancer or cardiovascular disease¹. Diseases that have long plagued humanity, such as malaria and tuberculosis, continue to impose a staggering toll. Recent decades have also witnessed the emergence of new virulent pathogens, including human immunodeficiency virus (HIV), Ebola virus, severe acute respiratory syndrome (SARS) coronavirus, West Nile virus and Zika virus.

The persistent global threat posed by microbial pathogens arises from the nonlinear mechanisms of disease transmission. That is, as the prevalence of a disease is reduced, the density of immune individuals drops, the density of susceptible individuals rises and disease is more likely to rebound. The resultant temporal trajectories are difficult to predict without considering this nonlinear interplay. For instance, many microbial diseases exhibit periodic spikes in the number of cases that are unexplainable by pathogen natural history or environmental phenomena. By explicitly defining the nonlinear processes underlying infectious disease spread, transmission models illuminate these otherwise opaque systems.

Forty years ago, *Nature* published a series of papers that launched the modern era of infectious disease modelling^{2,3}. Since that time, these methodologies have multiplied⁴. Transmission models now employ a variety of approaches, ranging from agent-based simulations that represent each individual⁵ to compartmental frameworks that group individuals by epidemiological status, such as infectiousness and immunity^{2,3}. Accompanying the methodological innovations, however, are challenges regarding selection of appropriate model structures from among the wealth of possibilities⁶.

At this anniversary of the publication of these landmark papers^{2,3}, we reflect on contributions that transmission modelling has made to infectious disease science and control. Through a series of case studies, we illustrate the overarching principles and challenges related to model design. With expanding computational capacity and new types of data, myriad opportunities have opened for transmission modelling to bolster evidence-based policy (Box 1)^{7,8}. In all

pursuits, modelling is most informative when conducted collaboratively with microbiologists, immunologists and epidemiologists. We offer this Perspective as an entry point for non-modelling scientists to understand the power and flexibility of modelling, and as a foundation for the transdisciplinary conversations that bolster the field.

Modelling across scales

Even within the same disease system, the ideal model design depends on the specifics of the questions asked. Here, we highlight a series of models focused on one of the defining infectious agents of our era: HIV. The virus has challenged science, medicine and public health at every scale, from its deft immune evasion to its death toll of more than 35 million over the last four decades⁹. We describe how clinical needs, research questions and data availability have shaped the design of HIV models across these scales. Unless otherwise indicated, the term 'HIV' is inclusive of both HIV-1 and HIV-2.

Within-host models. At a within-host scale (Table 1), models can be used to simulate cellular interactions, immunological responses and treatment pharmacokinetics¹⁰. In such simulations, viral dynamics are often modelled using a compartmental structure, with the growth of one population, such as circulating virions, dependent on the size of another population, such as infected cells. For example, a seminal within-host model fit to viral load data by Perelson et al.¹¹ revealed high turnover rates of HIV-1, counter to what was then the prevailing assumption that HIV-1 remained dormant during the asymptomatic 'latency' phase. The corollary to these high rates of viral turnover was that drug resistance would likely evolve rapidly under monotherapy. Further analyses of this model indicated that a combination of at least three drugs was necessary to maintain drug sensitivity¹². Once combination therapy did become available, extension of the Perelson et al. model demonstrated that the two-phase decline in viral load observed following treatment initiation was attributable to a reservoir of long-lived infected cells¹³. With this insight also came the realization that prolonged treatment would be necessary to suppress viral load. The incorporation of

¹Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA. ²Center for Infectious Disease Modeling and Analysis, Yale School of Public Health, New Haven, CT, USA. ³Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada. ⁴Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA. ⁵Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. ⁶Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. *e-mail: alison.galvani@yale.edu

Box 1 | Modelling infectious disease transmission for evidence-based policy

There are the three principal objectives of modelling, all of which can inform public health policy.

Predicting disease spread. Models can be used to estimate the infectiousness of a pathogen within a given population. A fundamental concept is that of R_0 , the basic reproduction number, which quantifies the number of infections that would result from a single index case in a susceptible population. R_0 governs the temporal trajectory of an outbreak and the scale of interventions required for its containment. Models may be used to infer R_0 as well as forecast changes in R_0 that could drive transitions in epidemic dynamics, such as the shift from sporadic outbreaks to sustained chains of transmission. Example: assessing real-time Zika risk in Texas⁹⁰.

Selecting among alternative control strategies. Simultaneous field trials of multiple infectious disease control options are often infeasible. Models can simulate a wide range of control strategies and thus optimize public health policies according to translational objectives and real-world constraints. Modelling can also extrapolate from the individual clinical outcomes of interventions or novel therapeutics to the population-level impacts. Extrapolating to the population level is essential to evaluate the indirect benefits of interventions, including a reduction in transmission, or unanticipated repercussions, such as evolution of resistance. Example: comparing antibiotic ‘cycling’ versus ‘mixing’ to minimize the evolution of antimicrobial resistance¹⁰⁷.

Hypothesis testing. It is often logistically or ethically infeasible to empirically test scientific hypotheses in the field or experimentally. Modelling can identify parsimonious explanations of observed phenomena, including complex outcomes that can arise from the nonlinear processes common in microbiological systems. Even simple models can be useful to help us understand dynamics that are common to many microbiological systems through identification of basic mechanisms that apply across a range of infections. By examining a new infectious agent through the lens of previously characterized systems, models provide insight into the ways that a particular microbial infection might follow or break from typical patterns. Example: investigating whether individual heterogeneity within social networks significantly impacts disease spread²².

stochasticity into this within-host framework allowed model fitting to ‘viral blips’—transient peaks in viral load, even under antiretroviral treatment¹⁴. Analysis of this data-driven stochastic model demonstrated that homeostatic proliferation maintained the infected cell reservoir and produced these viral blips, a finding that was later confirmed experimentally^{15,16}. The implication for clinical care was that intensified antiretroviral treatment would be unable to eliminate the latent reservoir of infected cells as had been hypothesized, sparing patients from potentially fruitless trials with such regimens.

Individual-based models. Whereas the unit of interest for within-host modelling is an infected cell, the analogous unit for individual-based models is an infected person (Table 1)^{5,17,18}. Individual-based models are often used to explore the interplay between disease transmission and individual-level risk factors, such as comorbidities, sexual behaviours and age. Such models are capable of incorporating data with individual-level granularity, including those regarding contact patterns, patient treatment cascades and clinical outcomes. Individual-based models are uniquely suited for






representing overlap in individual-level risk factors and translating the implications of this overlap for public health policy. For example, an individual-based model was recently used to demonstrate that the majority of HIV transmission among people who inject drugs in New York City is attributable to undiagnosed infections¹⁸. These modelling results underscore the urgency for the city to invest in more comprehensive screening and improved diagnostic practices.

Population models. Most commonly, models are created at the population scale, capturing the spread of a pathogen through a large group (Table 1). At this scale, compartmental models shift in focus from the pathogen to the host. Unlike individual-based models, compartmental models will aggregate individuals with a similar epidemiological status. For instance, the archetypal ‘S–I–R’ model separates the entire population of interest into one of three categories: S, susceptible to infection; I, infected and infectious; or R, recovered and protected¹⁹. In practice, most models will have additional compartments or stratification beyond this simple structure. Age stratification is essential when either the disease risk or the intervention is age-specific. As an example, an age-stratified multi-pathogen model demonstrated that schistosomiasis prevention targeted to Zimbabwean schoolchildren could cost-effectively reduce HIV acquisition later in life²⁰. This framework was extended to additional countries with a range of age-specific disease prevalence and co-infection rates to assess the potential value of treating schistosomiasis in adults. Although adult treatment is not usually considered efficient, the model showed that it could be cost-effective in settings with high HIV prevalence²¹. These models strengthened the investment case for treatment of schistosomiasis, an otherwise neglected tropical disease.

Network models are also deployed to represent dynamics on the population scale (Table 1). These models impose a structure on contacts between hosts, unlike compartmental models which assume that contacts are random among hosts within a compartment. In a network model, nodes represent individuals and the connections between nodes represent contacts through which infection may spread²². Sources for network parameterization may include surveys, partner notification services or phylogenetic tracing^{23,24}. As with individual-based models, network models tend to require significant amounts of data to fully parameterize, but various computational and statistical methods have been developed to analyse the impact of uncertain parameter values on model predictions²⁵. Network models are applied to discern the influence of contact structure on disease transmission and on the effectiveness of targeted intervention strategies. For instance, network models predicted that HIV would spread more quickly through sexual partnerships that are concurrent versus serially monogamous, even if the total numbers of sexual acts and partners remain constant²⁶. The study prompted a more rigorous engagement of epidemiologists with sociological data to tailor interventions for specific settings²⁷. Other network models have focused on the more rapid transmission within clusters of high-risk individuals and slower transmission to lower-risk clusters, a dynamic which explains discrepancies between observed incidence patterns and the expected pattern based on an assumption of homogeneous risks²⁸. These studies both illustrate the importance of accounting for network-driven dynamics when individuals are highly aggregated with regards to their risk factors, and when appropriate data for parameterization are available.

Metapopulation models. Metapopulation models represent disease transmission at dual scales, considering not just the interactions of individuals, but also the relationships between groups of individuals, which are typically defined geographically (Table 1). Transmission intensity is often higher within groups than across groups, especially when the groups are spatially segregated²⁹. One metapopulation model of HIV in mainland China considered

Table 1 | Scales of modelling for the HIV epidemic

	Scale	Potential data inputs	Model types	HIV examples
	Within-host	Viral dynamics; treatment dosage	C, IB	Refs. 11–14
	Individual	Access to care and treatment; co-morbidities; needle-sharing behaviour	IB	Refs. 17,18
	Population	Contact matrices; sexual partnership surveys; co-infection rates	C, N	Refs. 20,21,26–28
	Meta-population	Migration patterns	M	Ref. 30
	Global	Country-specific demographics, prevalence and drivers	C	Ref. 31

The research question and available data inputs drive the choice of model scale and type. At each scale, one or more model types may be appropriate, including compartmental (C), individual-based (IB), network (N) and metapopulation (M).

transmission within and between provinces, driven by the mobility of migrant labourers³⁰. The study suggested that HIV prevention resources could be most effectively targeted to provinces with the greatest initial incidence, as rising incidence in other provinces is driven more by migration from the high-burden provinces than by local transmission. Given that the Chinese provinces with employment opportunities for migrants are also those with the heaviest burden of HIV, migrant workers who acquire HIV often do so in the province where they work. However, government policy requires migrants to return to their home province for treatment. The movement of these workers perpetuates the disease cycle, as new migrants move to fill the vacated jobs and themselves become exposed to elevated HIV risk. These results therefore call for reconsideration of provincial treatment restrictions.

Multinational models. Global policies, such as the treatment goals set by the Joint United Nations Programme on HIV/AIDS (UNAIDS), have been modelled on a global scale (Table 1) by considering the effectiveness of the policies for each nation. For example, a compartmental model was used to evaluate the potential impact of a partially efficacious HIV vaccine on the epidemiological trajectories in 127 countries that together constitute over 99% of the global burden³¹. The model was tailored to each country by fitting to country-specific incidence trends as well as diagnosis, treatment and viral suppression data. This model revealed that, even with efficacy as low as 50%, a HIV vaccine would avert millions of new infections worldwide, irrespective of whether ambitious treatment goals are met. These results identify the synergies between vaccination and treatment-as-prevention, and provide evidence to support continued investment in vaccine development^{9,32}.

From the cellular level to the population level, HIV modelling has led to improvements in drug formulations, clinical care and resource allocation. As scientific advances continue to bring pharmaceutical innovations, modelling will remain a useful tool for illuminating transmission dynamics and optimizing public health policy.

Modelling emerging and re-emerging pathogens

HIV was not controlled before it became a pandemic, but our response to future outbreaks has the potential to be more timely³³. When diseases emerge in new settings, such as Ebola in West Africa and SARS in China, modelling can be rapidly deployed to inform

and support response efforts (Fig. 1). Unfortunately, the urgency of public health decisions during such outbreaks tends to be accompanied by a sparsity of data with which to parameterize, calibrate and validate models. As detailed below, uncertainty analysis—a method of analysing how uncertainty in input parameters translates to uncertainty in model outcome variables—becomes all the more vital in these situations. Media attention regarding model predictions is often heightened during outbreaks, ironically at a time when modelling results are apt to be less robust than for well-characterized endemic diseases. We discuss the importance of careful communication regarding model recommendations and associated uncertainty to inform the public without fuelling excessive alarm. Despite these challenges, and especially if these challenges can be navigated, the timely assessment of a wide range of intervention scenarios made possible by modelling would be particularly valuable during infectious disease emergencies.

Ebola virus outbreaks. The 2014 Ebola virus outbreak struck a populous region near the border of Guinea and Sierra Leone, sparking a crisis in a resource-constrained area that had no prior experience with the virus. As the caseload mounted and disseminated geographically, it became apparent that the West African outbreak would be unprecedented in its devastation. Models were developed to estimate the potential size of the epidemic in the absence of intervention, demonstrating the urgent need for expanded action by the international community^{34–36}, and to calculate the scale of the required investment³⁷. Initial control efforts included a militarily enforced quarantine of a Liberian neighbourhood in which Ebola was spreading. Modelling analysis in collaboration with the Liberian Ministry of Health demonstrated that the quarantine was ineffective and possibly even counterproductive³⁸. Connecting the microbiological and population scales, another modelling study integrated within-host viral load data over the course of Ebola infection and between-host transmission parameterized by contact-tracing data. The resulting dynamics highlighted the imperative to hospitalize most cases in isolation facilities within four days of symptom onset³⁹. These modelling predictions were borne out of empirical observations. Early in the outbreak, when the incidence was precipitously growing, the average time to hospitalization in Liberia was above six days⁴⁰. As contact tracing improved, the concomitant acceleration in hospitalization was found to be instrumental in turning the tide

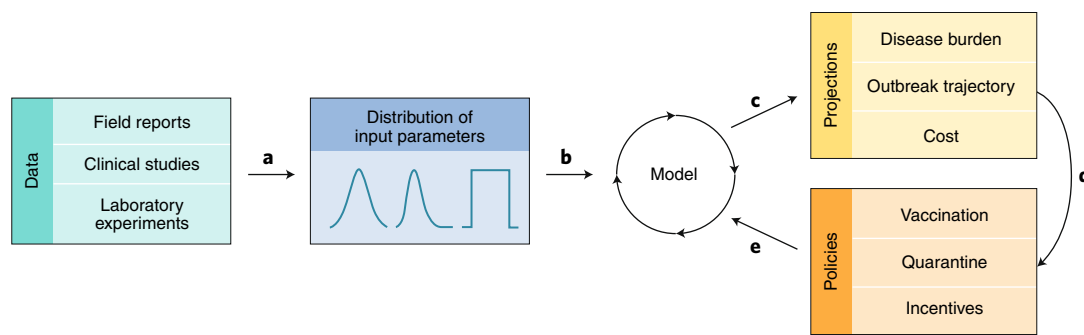


Fig. 1 | Data-driven model prediction to evaluate the impact of manipulatable policy variables. **a**, Data from a variety of sources, including surveillance reports, experiments and epidemiological studies, can inform model parameters. **b**, Rather than extracting single point estimates, modellers can use data more powerfully by constructing data-driven distributions for parameters from which values are sampled for each simulation. **c**, Every simulation yields a projection, such that multiple runs based on drawing probabilistically from empirical distributions generate a probabilistic distribution of projections. Types of projection that can be generated include outbreak trajectories, disease burdens and economic impact. **d**, Probabilistic uncertainty analyses convey not only model projections of policy outcomes, but also quantification of confidence in the projections. **e**, As policies are adopted and the microbiological system is influenced accordingly, the model can be iteratively updated to reflect the shifting status quo, thereby progressively optimizing policies within an evolving system.

on the outbreak⁴⁰. In another approach, phylogenetic analysis and transmission modelling were combined to estimate underreporting rates and social clustering of transmission⁴¹. This study informed public health authorities regarding the optimal scope and targeting of their efforts, which were central to stemming the epidemic.

Although data can be scarce for emerging pathogens, modellers can exploit similarities with better-characterized disease systems to investigate the potential efficiency of different interventions (Box 1). As vaccine candidates became available against Ebola, ring vaccination was proposed based on the success of the strategy in eliminating smallpox⁴², another microorganism whose transmission required close contact between individuals and for which peak infectiousness occurs after the appearance of symptoms. Compartmental models had suggested parameter combinations for which ring vaccination would be superior to mass vaccination⁴³, and methodological advances subsequently allowed for explicit incorporation of contact network data⁴⁴. Modelling based on social and healthcare contact networks specific to West Africa supported implementation of ring vaccination⁴⁵, and the approach was adopted for the clinical trial of the vaccine⁴⁶.

In 2018, two independent outbreaks of Ebola erupted in the Democratic Republic of the Congo. During the initial outbreak in Équateur province, modellers combined case reports with time series from previous outbreaks to generate projections of final epidemic size that could inform preparedness planning and allocation of resources⁴⁷. Ring vaccination was again deployed, this time within two weeks of detecting the outbreak. A spatial model quantified the impact of vaccine on both the ultimate burden and geographic spread of Ebola, highlighting how even one week of additional delay would have substantially reduced the ability of vaccination to contain this outbreak⁴⁸. The second outbreak was reported in August in the North Kivu province. Armed conflict in this region has interfered with the ability of healthcare workers to conduct the necessary contact tracing, vaccination and treatment. As conditions make routine data collection difficult and even dangerous, modelling has the potential to provide crucial insights into the otherwise unobservable characteristics of this outbreak.

Pandemic influenza. In contrast to the unexpected emergence of Ebola in a new setting, the influenza virus has repeatedly demonstrated its ability to cause pandemics. A pandemic is an event in which a pathogen creates epidemics across the entire globe. The 1918 pandemic killed an estimated 50 million people worldwide⁴⁹, exceeding the combined military and civilian casualties of World

War 1. While the 2% case-fatality rate of the 1918 strain was approximately 40 times higher than is typical for influenza⁵⁰, pathogenic strains with case-fatality rates exceeding 50% periodically emerge⁵¹. Modelling has illustrated how repeated zoonotic introductions impose selection for elevated human-to-human transmissibility, which thereby exacerbates the threat of a devastating influenza pandemic⁵². Such threats underscore the importance of surveillance systems and preparedness plans, which can be informed by modelling (Box 1). Transmission models are able to optimize surveillance systems, accelerate outbreak detection and improve forecasting^{53–56}. For example, a spatial model integrating a variety of surveillance data streams and embedded in a user-friendly platform is currently implemented by the Texas Department of State Health Services to generate real-time influenza forecasts (<http://flu.tacc.utexas.edu/>). Modelling has also motivated the development of dynamic preparedness plans, which adapt in response to the unfolding events of a pandemic, as models identified that adaptive efforts would be more likely to contain an influenza pandemic than static policies chosen a priori⁵⁷. Other pandemic influenza analyses used age-structured compartmental models to study the trade-off between targeting influenza vaccination to groups that transmit many infections but experience relatively low health burdens (for example, schoolchildren) versus groups that transmit fewer infections but experience greater health burdens (for example, the elderly)⁵⁸. Such examples illustrate the insights that modelling has provided to the decision makers charged with maintaining readiness against simultaneously rare but catastrophic situations.

Vaccine hesitancy. Modelling has also examined the impact of human behaviour, including vaccination decisions and social interactions, on the course of an epidemic. Public health interventions are not always sufficient to ensure disease control, as behavioural factors can thwart progress^{59–62}. For example, reports in 1974 of potential neurological side effects from the whole-cell pertussis vaccine led to a steep decline in vaccine uptake throughout the UK, followed by a slow recovery (Fig. 2a)⁶³. Vaccine uptake ebbed and flowed over the next two decades, with higher rates of vaccination in the wake of large pertussis outbreaks (Fig. 2b)^{61,63}. Compartmental models analysing the interplay between vaccine uptake and disease dynamics confirmed the hypothesis that increases in vaccination were a response to the pertussis infection risk⁶¹, and showed that incorporating this interplay can improve epidemiological forecasts. Network models extending these coupled disease-behaviour analyses

have illustrated how the perceived risk of vaccination can have greater influence on vaccine uptake than disease incidence⁶⁴.

More recently, vaccine refusal has led to the resurgence of measles in the USA^{62,65}. Researchers are turning to social media to gather information about attitudes toward vaccines and infectious diseases, and to glean clues about vaccinating behaviour^{55,66,67}. For instance, signals that vaccine refusal is compromising elimination can be detected months or years in advance of disease resurgence by applying mathematical analysis of tipping points to social media data that have been classified on the basis of sentiment using machine learning algorithms⁶⁶. These and other data science techniques might help public health authorities identify the specific communities that are at increased risk of future outbreaks. On shorter timescales, the near-instantaneous availability of social media data facilitates its integration into models developed for outbreak response^{55,66}. Other behavioural factors that have been incorporated into transmission models include attendance at social gatherings, sexual behaviour and commuting patterns—elements which are also often affected by perceived infection risk^{59,68,69}.

Antimicrobial resistance. A substantial portion of the increase in human lifespan over the last century is attributable to antibiotics⁷⁰, but the emergence of pathogen strains that are resistant to antimicrobials threatens to reverse these gains. The extensive use and misuse of antibiotics has led to the evolution of multidrug-resistant, extensively drug-resistant and even pan-drug-resistant pathogens across the globe. Precariously, this evolution outpaces the development of new antibiotics.

Mathematical modelling is being used to identify strategies to forestall the emergence and re-emergence of antimicrobial resistance^{71,72}. Models are particularly valuable for comparing alternative strategies, such as administration of different antibiotics within the same hospital ward, temporal cycling of antibiotics and combination therapy^{73–76}. High-performance computing now permits the rapid exploration of multidimensional parameter space. Models can thereby narrow an array of possible interventions down to a subset likely to have the highest impact or optimize between trade-offs, such as effectiveness and cost (Box 1). By contrast, expense, feasibility and ethical considerations may impose more limitations on in vivo investigations (Box 1). Not only can models identify the optimal strategy for a given parameter set, but they can generate the probability that this intervention remains optimal across variation in the parameters. For example, an optimization routine combined with simulation of hospital-based interventions identified combination therapy as most likely to reduce antibiotic resistance⁷⁵. As a complementary approach, modelling can incorporate economic considerations into these evaluations. A stochastic compartmental model showed that infection control specialists dedicated to promoting hand hygiene in hospitals are cost-effective for limiting the spread of antibiotic resistance⁷⁴.

Although most models of antibiotic resistance have focused on transmission in healthcare settings, the importance of antibiotic resistance in natural, agricultural and urban settings has been increasingly recognized^{77–83}. For example, a metapopulation model of antimicrobial-resistant *Clostridium difficile* simulated its transmission within and between hospitals, long-term care facilities and the community. This model demonstrated that mitigating risk in the community has the potential to substantially avert hospital-onset cases by decreasing the number of patients with colonization at admission and thereby the transmission within hospitals⁸⁴. This study illustrates how models can consider the entire ecosystem of infection to elucidate dynamics that might not be captured through focus on a single setting.

Sensitivity and uncertainty analysis

During the initial phase of an outbreak, the predictive power of models is often constrained by data scarcity. This challenge is

exacerbated for outbreaks of novel emerging diseases given that our understanding of the disease will rely on the unfolding epidemic (Fig. 1). Not only can the absence of data constrain model design, but sparse data requires extensive sensitivity analyses to evaluate the robustness of conclusions. Univariate sensitivity analyses, in which individual parameters are varied incrementally above and below a point estimate, can identify which parameters most influence model output (Box 1). Such comparisons reveal both salient gaps in knowledge and targets for preventing and mitigating the outbreak (Box 1)⁸⁵. As an outbreak progresses, each day has the potential to provide more information about the new disease, including its duration of latency, the symptomatic period, infectiousness, transmission modalities, underreporting and the case-fatality rate. However, collecting detailed data to inform each of these parameters can strain resources when they are thinly spread during an emergency response. Sensitivity analysis can support clinicians and epidemiologists in prioritizing data collection efforts⁸⁶.

Parameterization challenges are compounded for complicated disease systems, such as vector-borne diseases. For example, models of Zika virus infection span both species and scales, as the disease trajectory is influenced by factors ranging from mosquito seasonality and mosquito abundance down to viral and immunological dynamics within human and mosquito hosts^{87,88}. Adding to this complexity, the ecological parameters vary seasonally and geographically—heterogeneities that may be amplified by socioeconomic factors modulating human exposure to infected mosquitoes⁸⁹. In the absence of the high-resolution data that would be ideal to tailor a mosquito-driven disease system to a given setting, uncertainty analysis can unify parameterization from disparate data sources. In contrast to univariate sensitivity analyses, uncertainty analysis simultaneously samples from empirical- or expert-informed distributions for many or all input parameters. Collaboration between modellers and disease experts is thus instrumental to ensuring the biological plausibility of these parameter distributions^{90,91}. The uncertainty analysis produces both a central point estimate and a range for each outcome, a combination which can inform stakeholders about the best-case and worst-case scenarios as well as the likelihood that an intervention will be successful^{92–94}.

Pitfalls and how to avoid them

In constructing models and communicating results, there are common pitfalls which can compromise the rigor and impact of the research.

A pervasive pitfall is the incorporation of excessive model complexity, particularly through inclusion of more parameters than can be reliably parameterized from data. Intuition might suggest that a complex representation of a microbiological system would more closely represent reality. However, the predictive power of a model can be degraded if incorporating additional parameters only marginally improves the fit to data. This tendency results in complicated transmission models that overfit data in much the same way that complicated statistical regressions can overfit data, replicating not only the relevant trends but also the noise in a particular data set. These overfit models thus become less useful for prediction and generalization^{6,95}.

To guide appropriate model complexity and parameterization, modellers have used the mathematical theory of information to develop criteria which quantify the balance between realism and simplicity. Such criteria penalize additional parameters but reward substantial improvements in fit, thereby identifying the simplest model that can adequately fit the data^{61,96,97}. These methods can be applied to select among models or alternatively to calculate weighted average predictions across models. In a similar vein, modelling consortiums serve to address uncertainty surrounding model design^{98–100}. In a consortium, several modelling groups develop their models independently, each applying their particular

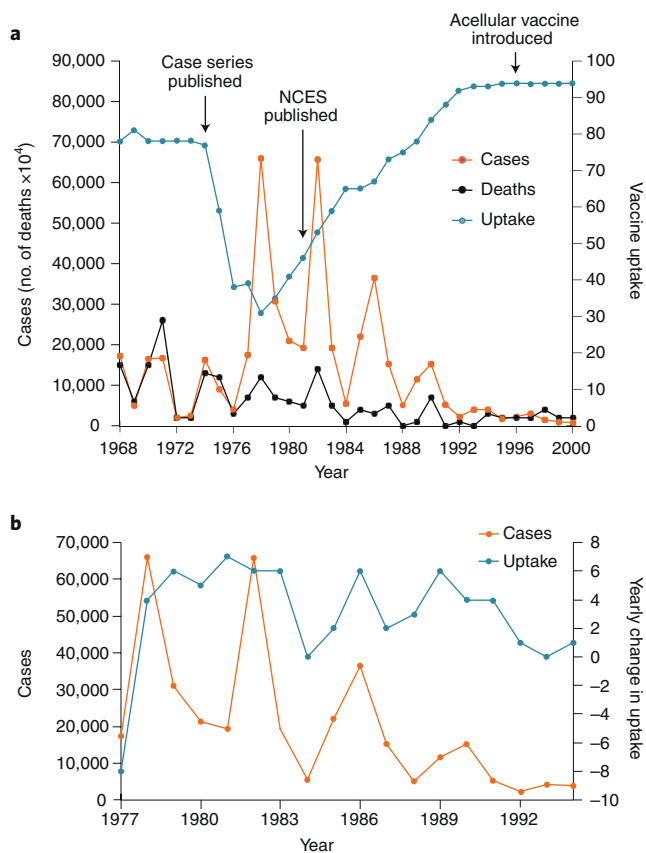


Fig. 2 | Behavioural changes drive outbreak patterns and also respond to them. **a**, Pertussis case notifications, pertussis deaths and the percentage of children completing the full course of vaccines by their second birthday in England and Wales, 1968–2000. A case series describing children with suspected neurological damage from the whole-cell pertussis vaccine was published in 1974 and received widespread media attention. Subsequently, the National Childhood Encephalopathy Study published in 1981 clarified the risks, which motivated public health efforts to boost vaccine uptake. The whole-cell pertussis vaccine was replaced with an acellular formulation in 1996. **b**, Pertussis case notifications and percentage change in vaccine uptake in successive years during the recovery phase, 1977–1994. Vaccine uptake appears to be entrained by surges in infection incidence. Mathematical models can capture the interplay between natural and human dynamics exemplified in this dataset and a wide variety of other study systems.

expertise and perspective. For example, consortia of malaria modellers were convened to predict the effectiveness of interventions, including a vaccine candidate¹⁰¹ and mass drug administration¹⁰². Congruence of output among models engenders confidence that model results are robust.

Another pitfall concerns the quality of data used to inform the model. Incompleteness of data has been an issue since 1766, when Daniel Bernoulli published a compartmental model of smallpox and acknowledged that more extended analyses would have been possible if the data had been age-stratified¹⁰³. Even today, using data to develop models without knowledge of how the data were collected or the limitations of the data can be risky. Data collected for an alternative purpose can contain gaps or biases that are acceptable for the original research question, yet lead to incorrect conclusions when incorporated for another purpose in a specific model. In ideal circumstances, modellers would be involved in the design of the original study, ensuring both seamless integration of the results into the model and awareness on the part of the modeller with regard to

data limitations. Failing that, it is very helpful for modellers to collaborate with scientists familiar with the details of empirical studies on which their results might depend.

This lack of familiarity with the biases or incompleteness of data sources may be particularly dangerous in the era of digital data. ‘Big data hubris’ can blind researchers to the limitations of the dataset, such as being a large but unrepresentative sample of the general population, or the alteration of search engine algorithms partway through the data collection process⁶. Some of these limitations can be addressed by using digital data as a complement to traditional data sources. In this way, the weakness of one data source (for example, low sample size of traditional surveys or bias in large digital data) can be compensated by the strengths of another data source (for example, balanced representation in small survey versus large scale of digital data).

A final pitfall that often arises in the midst of an ongoing outbreak concerns the interpretation of epidemic projections. Initial models may assume an absence of intervention as a way to assess the potential consequences of inaction. Such projections may contribute to the mobilization of government resources towards control, as was the case during the West African Ebola outbreak^{35,37,38}. In this respect, the projections are intended to make themselves obsolete¹⁰⁴. In retrospect and without knowledge of the initial purpose of the model, it may appear that the initial predictions were excessively pessimistic¹⁰⁵. Additionally, people living in outbreak zones often change their behaviour to reduce infection risks, thereby mitigating disease spread through, for example, reducing social interactions or increasing vaccine uptake (Fig. 2)^{59,61,66}. Thus, risk assessment constitutes a ‘moving target’¹⁰⁵. For example, input parameters estimated from contact tracing early in an outbreak could require adjustments to reflect these behaviour changes and accurately predict subsequent dynamics¹⁰⁶.

The need for proficient communication skills is heightened during an outbreak. This concern is particularly relevant when presenting sensitivity and uncertainty analyses. Although predictions at the extreme of sensitivity analyses also tend to be less probable than mid-range projections, there can be a temptation to focus on the most sensational model scenarios. Ensuing public pressure on the basis of misunderstood findings can cause unwarranted alarm and trigger counterproductive political decisions. In both publications and media interactions, underscoring the improbability of extreme scenarios explored during sensitivity analysis, as well as how improved interventions turn a predictive model into a counterfactual one, may pre-empt this pitfall³³.

Conclusion

The role for modelling in supporting epidemiologists, public health officials and microbiologists has progressively expanded since the foundational publications forty years ago, in concert with the growing abundance and granularity of data as well as the refinement of quantitative approaches. Models have now been developed for virtually every human infectious disease, as well as in many that affect animals and plants, and have been applied across the globe. Interdisciplinary collaboration among empiricists, policymakers and modellers facilitates the development of scientifically grounded models for specific settings and generates results that will be actionable in the real world. Reciprocally, modelling results may guide the design of experiments and field studies by revealing key gaps in our understanding of microbiological systems. Furthermore, modelling is a feasible and cost-effective approach for identifying impactful policies prior to implementation decisions. Through all these avenues, epidemiological modelling galvanizes evidence-based action to alleviate disease burden and improve global health.

Received: 24 April 2018; Accepted: 15 August 2019;
Published online: 20 September 2019

References

- Naghavi, M. et al. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* **390**, 1151–1210 (2017).
- Anderson, R. M. & May, R. M. Population biology of infectious diseases: part I. *Nature* **280**, 361–367 (1979).
- May, R. M. & Anderson, R. M. Population biology of infectious diseases: part II. *Nature* **280**, 455–461 (1979).
- Heesterbeek, H. et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, aaa4339 (2015).
- Marshall, B. D. L. & Galea, S. Formalizing the role of agent-based modeling in causal inference and epidemiology. *Am. J. Epidemiol.* **181**, 92–99 (2015).
- Lazer, D., Kennedy, R., King, G. & Vespignani, A. Big data. The parable of Google Flu: traps in big data analysis. *Science* **343**, 1203–1205 (2014).
- Walls, H. L., Peeters, A., Reid, C. M., Liew, D. & McNeil, J. J. Predicting the effectiveness of prevention: a role for epidemiological modeling. *J. Prim. Prev.* **29**, 295–305 (2008).
- Knight, G. M. et al. Bridging the gap between evidence and policy for infectious diseases: how models can aid public health decision-making. *Int. J. Infect. Dis.* **42**, 17–23 (2016).
- Corey, L. & Gray, G. E. Preventing acquisition of HIV is the only path to an AIDS-free generation. *Proc. Natl Acad. Sci. USA* **114**, 3798–3800 (2017).
- Cappuccio, A., Tieri, P. & Castiglione, F. Multiscale modelling in immunology: a review. *Brief. Bioinform.* **17**, 408–418 (2016).
- Perelson, A. S., Neumann, A. U., Markowitz, M., Leonard, J. M. & Ho, D. D. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* **271**, 1582–1586 (1996).
- Perelson, A. S., Essunger, P. & Ho, D. D. Dynamics of HIV-1 and CD4⁺ lymphocytes in vivo. *AIDS* **11**(Suppl. A), S17–S24 (1997).
- Perelson, A. S. et al. Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188–191 (1997).
- Rong, L. & Perelson, A. S. Modeling latently infected cell activation: viral and latent reservoir persistence, and viral blips in HIV-infected patients on potent therapy. *PLoS Comput. Biol.* **5**, e1000533 (2009).
- Chomont, N. et al. HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation. *Nat. Med.* **15**, 893–900 (2009).
- Perelson, A. S. & Ribeiro, R. M. Modeling the within-host dynamics of HIV infection. *BMC Biol.* **11**, 96 (2013).
- Eaton, J. W. et al. Assessment of epidemic projections using recent HIV survey data in South Africa: a validation analysis of ten mathematical models of HIV epidemiology in the antiretroviral therapy era. *Lancet Glob. Health* **3**, e598–e608 (2015).
- Escudero, D. J. et al. The risk of HIV transmission at each step of the HIV care continuum among people who inject drugs: a modeling study. *BMC Public Health* **17**, 614 (2017).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, 1991).
- Ndeffo Mbah, M. L. et al. Cost-effectiveness of a community-based intervention for reducing the transmission of *Schistosoma haematobium* and HIV in Africa. *Proc. Natl Acad. Sci. USA* **110**, 7952–7957 (2013).
- Ndeffo Mbah, M. L., Gilbert, J. A. & Galvani, A. P. Evaluating the potential impact of mass praziquantel administration for HIV prevention in *Schistosoma haematobium* high-risk communities. *Epidemics* **7**, 22–27 (2014).
- Bansal, S., Grenfell, B. T. & Meyers, L. A. When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* **4**, 879–891 (2007).
- Delva, W., Leventhal, G. E. & HELLERINGER, S. Connecting the dots: network data and models in HIV epidemiology. *AIDS* **30**, 2009–2020 (2016).
- Campbell, E. M. et al. Detailed transmission network analysis of a large opiate-driven outbreak of HIV infection in the United States. *J. Infect. Dis.* **216**, 1053–1062 (2017).
- Dutta, R., Mira, A. & Onnela, J.-P. Bayesian inference of spreading processes on networks. *Proc. Math. Phys. Eng. Sci.* **474**, 20180129 (2018).
- Morris, M. & Kretzschmar, M. Concurrent partnerships and the spread of HIV. *AIDS* **11**, 641–648 (1997).
- Kretzschmar, M., White, R. G. & Caraël, M. Concurrency is more complex than it seems. *AIDS* **24**, 313–315 (2010).
- Friedman, S. R. et al. Network-related mechanisms may help explain long-term HIV-1 seroprevalence levels that remain high but do not approach population-group saturation. *Am. J. Epidemiol.* **152**, 913–922 (2000).
- Beyer, H. L. et al. Metapopulation dynamics of rabies and the efficacy of vaccination. *Proc. Biol. Sci.* **278**, 2182–2190 (2011).
- Xiao, Y. et al. Predicting the HIV/AIDS epidemic and measuring the effect of mobility in mainland China. *J. Theor. Biol.* **317**, 271–285 (2013).
- Medlock, J. et al. Effectiveness of UNAIDS targets and HIV vaccination across 127 countries. *Proc. Natl Acad. Sci. USA* **114**, 4017–4022 (2017).
- Fauci, A. S. An HIV vaccine is essential for ending the HIV/AIDS pandemic. *JAMA-J. Am. Med. Assoc.* **318**, 1535–1536 (2017).
- Lofgren, E. T. et al. Opinion: mathematical models: a key tool for outbreak response. *Proc. Natl Acad. Sci. USA* **111**, 18095–18096 (2014).
- WHO Ebola Response Team. et al. Ebola virus disease in West Africa—the first 9 months of the epidemic and forward projections. *N. Engl. J. Med.* **371**, 1481–1495 (2014).
- Meltzer, M. I. et al. Estimating the future number of cases in the Ebola epidemic — Liberia and Sierra Leone, 2014–2015. *MMWR-Morb. Mortal. W.* **63**, 1–14 (2014).
- Townsend, J. P., Skrip, L. A. & Galvani, A. P. Impact of bed capacity on spatiotemporal shifts in Ebola transmission. *Proc. Natl Acad. Sci. USA* **112**, 14125–14126 (2015).
- Lewnard, J. A. et al. Dynamics and control of Ebola virus transmission in Montserrado, Liberia: a mathematical modelling analysis. *Lancet Infect. Dis.* **14**, 1189–1195 (2014).
- Pandey, A. et al. Strategies for containing Ebola in West Africa. *Science* **346**, 991–995 (2014).
- Yamin, D. et al. Effect of Ebola progression on transmission and control in Liberia. *Ann. Intern. Med.* **162**, 11–17 (2015).
- Fallah, M. et al. Interrupting Ebola transmission in Liberia through community-based initiatives. *Ann. Intern. Med.* **164**, 367–369 (2016).
- Scarpino, S. V. et al. Epidemiological and viral genomic sequence analysis of the 2014 Ebola outbreak reveals clustered transmission. *Clin. Infect. Dis.* **60**, 1079–1082 (2014).
- Foege, W. H., Millar, J. D. & Lane, J. M. Selective epidemiologic control in smallpox eradication. *Am. J. Epidemiol.* **94**, 311–315 (1971).
- Kaplan, E. H., Craft, D. L. & Wein, L. M. Emergency response to a smallpox attack: the case for mass vaccination. *Proc. Natl Acad. Sci. USA* **99**, 10935–10940 (2002).
- House, T. & Keeling, M. J. The impact of contact tracing in clustered populations. *PLoS Comput. Biol.* **6**, e1000721 (2010).
- Wells, C. et al. Harnessing case isolation and ring vaccination to control Ebola. *PLoS Negl. Trop. Dis.* **9**, e0003794 (2015).
- Henao-Restrepo, A. M. et al. Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!). *Lancet* **389**, 505–518 (2017).
- Kelly, J. D. et al. Projections of Ebola outbreak size and duration with and without vaccine use in Équateur, Democratic Republic of Congo, as of May 27, 2018. *PLoS ONE* **14**, e0213190 (2019).
- Wells, C. R. et al. Ebola vaccination in the Democratic Republic of the Congo. *Proc. Natl Acad. Sci. USA* **116**, 10178–10183 (2019).
- Neumann, G., Noda, T. & Kawaoka, Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature* **459**, 931–939 (2009).
- U.S. Department of Health & Human Services. Estimated influenza illnesses, medical visits, hospitalizations, and deaths averted by vaccination in the United States. *Centers for Disease Control and Prevention* https://www.cdc.gov/flu/about/burden-averted/2015-16.htm?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fflu%2Fabout%2F%2F2015-16.htm (2017).
- Lai, S. et al. Global epidemiology of avian influenza A H5N1 virus infection in humans, 1997–2015: a systematic review of individual case data. *Lancet Infect. Dis.* **16**, e108–e118 (2016).
- Antia, R., Regoes, R. R., Koella, J. C. & Bergstrom, C. T. The role of evolution in the emergence of infectious diseases. *Nature* **426**, 658–661 (2003).
- Brownstein, J. S. et al. Information technology and global surveillance of cases of 2009 H1N1 influenza. *N. Engl. J. Med.* **362**, 1731–1735 (2010).
- Scarpino, S. V., Dimitrov, N. B. & Meyers, L. A. Optimizing provider recruitment for influenza surveillance networks. *PLoS Comput. Biol.* **8**, e1002472 (2012).
- Salathé, M., Freifeld, C. C., Mekaru, S. R., Tomasulo, A. F. & Brownstein, J. S. Influenza A (H7N9) and the importance of digital epidemiology. *N. Engl. J. Med.* **369**, 401–404 (2013).
- Herrera, J. L., Srinivasan, R., Brownstein, J. S., Galvani, A. P. & Meyers, L. A. Disease surveillance on complex social networks. *PLoS Comput. Biol.* **12**, e1004928 (2016).
- Wallinga, J., van Boven, M. & Lipsitch, M. Optimizing infectious disease interventions during an emerging epidemic. *Proc. Natl Acad. Sci. USA* **107**, 923–928 (2010).
- Medlock, J. & Galvani, A. P. Optimizing influenza vaccine distribution. *Science* **325**, 1705–1708 (2009).
- Bauch, C. T. & Galvani, A. P. Epidemiology. Social factors in epidemiology. *Science* **342**, 47–49 (2013).
- Funk, S. et al. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics* **10**, 21–25 (2015).
- Bauch, C. T. & Bhattacharyya, S. Evolutionary game theory and social learning can determine how vaccine scares unfold. *PLoS Comput. Biol.* **8**, e1002452 (2012).
- U.S. Department of Health & Human Services. Measles cases and outbreaks. *Centers for Disease Control and Prevention* <https://www.cdc.gov/measles/cases-outbreaks.html> (2019).

63. Baker, J. P. The pertussis vaccine controversy in Great Britain, 1974–1986. *Vaccine* **21**, 4003–4010 (2003).
64. Bhattacharyya, S., Vutha, A. & Bauch, C. T. The impact of rare but severe vaccine adverse events on behaviour-disease dynamics: a network model. *Sci. Rep.* **9**, 7164 (2019).
65. Otterman, S. & Piccoli, S. Measles outbreak: opposition to vaccine extends well beyond ultra-Orthodox Jews in N. Y. *New York Times* <https://www.nytimes.com/2019/05/09/nyregion/measles-outbreak-ny-schools.html> (2019).
66. Pananos, A. D. et al. Critical dynamics in population vaccinating behavior. *Proc. Natl Acad. Sci. USA* **114**, 13762–13767 (2017).
67. Salathé, M. et al. Digital epidemiology. *PLoS Comput. Biol.* **8**, e1002616 (2012).
68. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl Acad. Sci. USA* **106**, 21484–21489 (2009).
69. Datta, S., Mercer, C. H. & Keeling, M. J. Capturing sexual contact patterns in modelling the spread of sexually transmitted infections: evidence using Natsal-3. *PLoS ONE* **13**, e0206501 (2018).
70. Fauci, A. S. & Marston, L. D. The perpetual challenge of antimicrobial resistance. *JAMA-J. Am. Med. Assoc.* **311**, 1853–1854 (2014).
71. Johnsen, P. J. et al. Factors affecting the reversal of antimicrobial-drug resistance. *Lancet Infect. Dis.* **9**, 357–364 (2009).
72. Johnsen, P. J. et al. Retrospective evidence for a biological cost of vancomycin resistance determinants in the absence of glycopeptide selective pressures. *J. Antimicrob. Chemother.* **66**, 608–610 (2011).
73. Ahmad, A. et al. Multistrain models predict sequential multidrug treatment strategies to result in less antimicrobial resistance than combination treatment. *BMC Microbiol.* **16**, 118 (2016).
74. Kardaś-Słoma, L. et al. Universal or targeted approach to prevent the transmission of extended-spectrum beta-lactamase-producing Enterobacteriaceae in intensive care units: a cost-effectiveness analysis. *BMJ Open* **7**, e017402 (2017).
75. Tepekule, B., Uecker, H., Derungs, I., Frenoy, A. & Bonhoeffer, S. Modeling antibiotic treatment in hospitals: a systematic approach shows benefits of combination therapy over cycling, mixing, and mono-drug therapies. *PLoS Comput. Biol.* **13**, e1005745 (2017).
76. van Kleef, E., Luangsanatip, N., Bonten, M. J. & Cooper, B. S. Why sensitive bacteria are resistant to hospital infection control. *Wellcome Open Res.* **2**, 16 (2017).
77. Allen, H. K. et al. Call of the wild: antibiotic resistance genes in natural environments. *Nat. Rev. Microbiol.* **8**, 251–259 (2010).
78. Carter, D. L. et al. Antibiotic resistant bacteria are widespread in songbirds across rural and urban environments. *Sci. Total Environ.* **627**, 1234–1241 (2018).
79. Xiang, Q. et al. Spatial and temporal distribution of antibiotic resistomes in a peri-urban area is associated significantly with anthropogenic activities. *Environ. Pollut.* **235**, 525–533 (2018).
80. Bueno, I. et al. Impact of point sources on antibiotic resistance genes in the natural environment: a systematic review of the evidence. *Anim. Health Res. Rev.* **18**, 1–16 (2017).
81. Szekeres, E. et al. Investigating antibiotics, antibiotic resistance genes, and microbial contaminants in groundwater in relation to the proximity of urban areas. *Environ. Pollut.* **236**, 734–744 (2018).
82. Bueno, I. et al. Systematic review: impact of point sources on antibiotic-resistant bacteria in the natural environment. *Zoonoses Public Health* **65**, e162–e184 (2018).
83. Bengtsson-Palme, J., Kristiansson, E. & Larsson, D. G. J. Environmental factors influencing the development and spread of antibiotic resistance. *FEMS Microbiol. Rev.* **42**, fux053 (2018).
84. Durham, D. P., Olsen, M. A., Dubberke, E. R., Galvani, A. P. & Townsend, J. P. Quantifying transmission of *Clostridium difficile* within and outside healthcare settings. *Emerg. Infect. Dis.* **22**, 608–616 (2016).
85. Barton, G. R., Briggs, A. H. & Fenwick, E. A. L. Optimal cost-effectiveness decisions: the role of the cost-effectiveness acceptability curve (CEAC), the cost-effectiveness acceptability frontier (CEAF), and the expected value of perfect information (EVPI). *Value Health* **11**, 886–897 (2008).
86. Gilbert, J. A., Meyers, L. A., Galvani, A. P. & Townsend, J. P. Probabilistic uncertainty analysis of epidemiological modeling to guide public health intervention policy. *Epidemics* **6**, 37–45 (2014).
87. Osuna, C. E. et al. Zika viral dynamics and shedding in rhesus and cynomolgus macaques. *Nat. Med.* **22**, 1448–1455 (2016).
88. Durham, D. P. et al. Evaluating vaccination strategies for Zika virus in the Americas. *Ann. Intern. Med.* **168**, 621–630 (2018).
89. Ramos, M. M. et al. Epidemic dengue and dengue hemorrhagic fever at the Texas–Mexico border: results of a household-based seroepidemiologic survey, December 2005. *Am. J. Trop. Med. Hyg.* **78**, 364–369 (2008).
90. Castro, L. A. et al. Assessing real-time Zika risk in the United States. *BMC Infect. Dis.* **17**, 284 (2017).
91. Fitzpatrick, M. C. et al. One Health approach to cost-effective rabies control in India. *Proc. Natl Acad. Sci. USA* **113**, 14574–14581 (2016).
92. Fitzpatrick, M. C. et al. Cost-effectiveness of canine vaccination to prevent human rabies in rural Tanzania. *Ann. Intern. Med.* **160**, 91–100 (2014).
93. Fitzpatrick, M. C. et al. Cost-effectiveness of next-generation vaccines: the case of pertussis. *Vaccine* **34**, 3405–3411 (2016).
94. Sah, P., Medlock, J., Fitzpatrick, M. C., Singer, B. H. & Galvani, A. P. Optimizing the impact of low-efficacy influenza vaccines. *Proc. Natl Acad. Sci. USA* **115**, 5151–5156 (2018).
95. Olson, D. R., Konty, K. J., Paladini, M., Viboud, C. & Simonsen, L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput. Biol.* **9**, e1003256 (2013).
96. Akaike, H. A new look at the statistical model identification. *IEEE T. Autom. Contr.* **19**, 716–723 (1974).
97. Johnson, J. B. & Omland, K. S. Model selection in ecology and evolution. *Trends Ecol. Evol.* **19**, 101–108 (2004).
98. Pitzer, V. E. et al. Direct and indirect effects of rotavirus vaccination: comparing predictions from transmission dynamic models. *PLoS ONE* **7**, e42320 (2012).
99. Rock, K. S. et al. Data-driven models to predict the elimination of sleeping sickness in former Equateur province of DRC. *Epidemics* **18**, 101–112 (2017).
100. Hollingsworth, T. D. & Medley, G. F. Learning from multi-model comparisons: collaboration leads to insights, but limitations remain. *Epidemics* **18**, 1–3 (2017).
101. Penny, M. A. et al. Public health impact and cost-effectiveness of the RTS, S/AS01 malaria vaccine: a systematic comparison of predictions from four mathematical models. *Lancet* **387**, 367–375 (2016).
102. Brady, O. J. et al. Role of mass drug administration in elimination of *Plasmodium falciparum* malaria: a consensus modelling study. *Lancet Glob. Health* **5**, e680–e687 (2017).
103. Dietz, K. & Heesterbeek, J. A. P. Daniel Bernoulli's epidemiological model revisited. *Math. Biosci.* **180**, 1–21 (2002).
104. Rivers, C. Ebola: models do more than forecast. *Nature* **515**, 492 (2014).
105. Butler, D. Models overestimate Ebola cases. *Nature* **515**, 18 (2014).
106. Epstein, J. M., Parker, J., Cummings, D. & Hammond, R. A. Coupled contagion dynamics of fear and disease: mathematical and computational explorations. *PLoS ONE* **3**, e3955 (2008).
107. Bergstrom, C. T., Lo, M. & Lipsitch, M. Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance in hospitals. *Proc. Natl Acad. Sci. USA* **101**, 13285–13290 (2004).

Acknowledgements

The authors gratefully acknowledge funding from the Notsew Orm Sands Foundation (grants to M.C.F., J.P.T. and A.P.G.), the National Institutes of Health (grant nos. K01 AI141576 and U01 GM087719 to M.C.F. and A.P.G., respectively) and the Natural Sciences and Engineering Research Council of Canada (grant no. RGPIN-04210-2014 to C.T.B.). The authors also thank C. Wells and A. Pandey, both members of the Yale Center for Infectious Disease Modeling and Analysis, for their helpful discussions regarding the HIV and Ebola modelling literature.

Author contributions

M.C.F. and A.P.G. drafted the initial manuscript. M.C.F., C.T.B., J.P.T. and A.P.G. all critically revised the content.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to A.P.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2019