



Modeling Approaches Toward Understanding Infectious Disease Transmission

14

Laura A. Skrip and Jeffrey P. Townsend

14.1 Introduction

The global burden of morbidity and mortality associated with communicable diseases is nearly three times that of cancer and 70% higher than that of cardiovascular disease [1]. Much of the infectious disease burden is caused by diseases that have a long history in human populations, such as measles, influenza, and pertussis. In recent years, additional disease burden has emerged from outbreaks of HIV, Chikungunya, Zika virus, *Ebola virus*, Middle East respiratory syndrome, coronavirus, *Nipah virus*, and Lassa fever, underscoring the continued vulnerability of human populations to novel and re-emerging threats. Mathematical models of infectious disease can be used to understand and predict epidemiological trajectories of outbreaks; they can then be applied to optimize strategies for mitigating disease transmission. These models incorporate pertinent environmental factors, such as seasonal fluctuations, that affect survival or activity of pathogens and their vectors. They can also address complexities of host behavioral change, including local and global human societal changes, health policy, clinical trial design, and resource allocation, which have been increasingly guided by modeling studies.

The epidemiology of infectious diseases and their control is complex. While models can be structured to represent the intricacies of a disease system, the quality of the information they provide depends on balancing existing data, methodological rigor, and simplifying assumptions. Here we describe factors influencing disease transmission, including both mechanisms of microbial infection and host behavior. We then discuss how model structure and parameterization are optimally used for addressing research and policy questions in the context of available information.

14.2 Biological Aspects of Disease Transmission

The spread and control of infectious diseases are related to biological factors that impact the magnitude, timing, and route of transmission.

L. A. Skrip

Department of Epidemiology of Microbial Disease, Yale School of Public Health, New Haven, CT, USA

J. P. Townsend (✉)

Department of Biostatistics, Yale School of Public Health, Yale University, New Haven, CT, USA

e-mail: jeffrey.townsend@yale.edu

14.2.1 Mode of Transmission

Transmission could result from direct contact between hosts, could require host-environment interaction, or could be mediated by a vector—often a phlebotomine insect—that transmits a pathogen between different hosts. Transmission can also vary across strata of the host population, as subgroups determined by behavioral, genetic, demographic, environmental, or other characteristics may result in differential exposure or susceptibility to specific diseases. Much of the art and science of modeling rides on critically evaluating which factors are essential to include in order to generate quantitative understanding of a focal research question.

14.2.2 Course of Infection and Infectiousness

Beyond accounting for the change in number of infected individuals over time, it can be important to also incorporate the course of infection during which microbial load, symptom severity, and contact behavior may all vary. Infection progression, or “age of infection,” is associated with temporal changes in infectiousness. Microbiological studies that measure the change in pathogen load over the course of an infection are integral to parameterization of dynamic models of disease, particularly evaluating interventions as the timing of intervention during the infection period could impact predicted effectiveness.

14.3 Behavior and Disease Transmission

The behaviors of human hosts, reservoir hosts, and vectors are fundamental to risks of both infection and transmission to others. For nonhuman hosts and vectors, the geographical distribution of habitat, feeding preferences, and proximity to humans affect the rate of transmission. For human hosts, contact patterns, belief systems, attitudes about public health recommendations, and propensity to seek care impact the real-world effectiveness of infection control measures, including quarantining, school closures, handwashing, and vaccination. To some extent, behavioral effects are implicit in the measures of transmission that are used in and estimated by models. However, in many cases it can be more appropriate to treat the transmission rate as a dynamic variable that evolves in concert with behaviors associated with the epidemic. This approach has been increasingly adopted when changes in willingness to accept interventions, in public health policy, or in disease awareness are expected to impact a component (i.e., contact rate, probability of transmission given contact, or frequencies of susceptible and infectious individuals) of overall transmission [2–6].

14.3.1 Willingness to Accept Interventions

Interventions, such as vaccination, have been instrumental in reducing the prevalence and incidence of many infectious diseases, particularly those affecting children, but they are only as effective as the extent to which individuals are willing to accept them. Vaccine scares exemplify the mutual feedback between host behavior and disease dynamics: low vaccine coverage can increase the probability of disease outbreaks. As disease outbreaks occur, people can become increasingly eager to vaccinate, which in turn reduces transmission. Thus, changes in incidence are mediated by behavior-based changes in the susceptibility of population members [5].

A challenge to tackling vaccine scares and other extreme reactions is that the human mind typically overestimates the risk of highly improbable events when the thought of such events evokes fear and anxiety [7]. Human risk perception is based on potential threats to one's immediate community—historically, populations of hundreds in geographically isolated areas. However, with the shift to a more global community, our intuitive understanding of risk has adapted in ways that foster overreaction. Increased awareness of international political and health crises and a sense of connectedness to regions simply an airplane flight away have led to a “problem of scale.” Teasing apart the influence of such factors from individual risk perception will be critical to determining how a model can represent elimination efforts within a given population.

14.3.2 Public Health Policy and Disease Awareness

In most cases of emerging or neglected diseases, vaccines are not available. Public health authorities must rely on other infection control methods, such as isolation of infected cases, quarantine of potentially exposed individuals, vector control, and/or social mobilization campaigns to enhance behavior change. Some authorities may enact recommendations and policies that increase social distancing, by which gatherings of individuals in affected areas are reduced through school or workplace closures. One approach to capturing how social distancing approaches reduce transmission is to model the social network explicitly, as opposed to assuming homogeneous mixing within the population. For instance, school networks could be superimposed on a dynamic model, keeping track of school-age individuals in the population and calculating their risks of infection as a function of transmission in both schools and the wider community. Even in the absence of enforced policies, individuals often spontaneously reduce their contacts in response to an outbreak by keeping children home from school, washing their hands more frequently, or avoiding public transit [8]. Such behavior change can be represented in a network model by incorporating a contact rate between network neighbors that depends on local rather than overall disease prevalence in the network [6].

Host behavioral factors are linked with epidemiological characteristics like mode of transmission. In the case of potentially airborne-transmitted and highly infectious measles, vaccination is one of the most effective interventions; therefore, population concern about vaccine safety must be addressed. In contrast, no vaccine was available in the initial phase of the 2014–2015 *Ebola* epidemic; because *Ebola virus* can easily be transmitted through close contact with the deceased, population attitudes toward funerary customs had to be addressed. Such examples also illustrate the mutual feedback that operates between host behavior and disease dynamics, wherein human behavior impacts transmission (e.g., reduced infection prevalence due to greater uptake of vaccines or participation in public health recommendations), but epidemiological trajectories can also influence behavior as perceptions of infection risk evolve. Such coupled behavior-disease systems can be represented by embedding dynamic transmission models within game theoretic or other behavioral frameworks. The resulting models can be informed by data regarding contact patterns and decision-making, as well as how the incubation period and the mode of transmission interact with host risk perception.

14.4 Infectious Disease Model Frameworks

Traditional biostatistical approaches (e.g., using regression techniques to identify predictor variables for specific epidemiological outcomes) focus on identifying associations between variables that may implicate causal relationships between them. In contrast, mathematical models of infectious disease are constructed by describing mechanistic processes that can include transmission between hosts,

immunological responses to pathogens within hosts, and vector ecological dynamics, among other dynamics. Simulation of the model then demonstrates how assumptions and existing information about the disease system relate to transmission dynamics.

14.4.1 Model Structures

There are myriad model frameworks that can be employed to account for host (e.g., age, sex, immunodeficiency) and environmental (e.g., climate conditions, prevalence of disease-carrying and disease-transmitting insects) factors of disease transmission in order to answer diverse epidemiological questions (Box 14.1, Fig. 14.1). Consequently, the ideal model structure for addressing particular research questions depends on the pertinent biological aspects of disease.

Box 14.1 Recent Methodological Innovations at the Frontier of Disease Transmission Modeling

Recent methodological innovations have expanded the applicability of disease transmission modeling as a tool to inform scientists and policy-makers before they undertake field research, laboratory experimentation, or program implementation.

1. Improvements in the rigor of field study and trial design

Epidemiological modeling techniques can be applied to simulate different trial designs and conduct power analyses prior to implementation. With the rapid development of vaccines against *Ebola virus* disease during the 2014–2015 outbreak in West Africa, field testing of the vaccines in high transmission settings was undertaken to simultaneously assess efficacy and prevent new cases among at-risk contacts. Modeling analyses determined false-positive rates and power across ranges of *Ebola* vaccine efficacy and trial start dates, comparing the stepped wedge and randomized controlled trial (RCT) designs [13]. Modeling revealed that the RCT design could achieve higher power for a given vaccine efficacy, largely due to the spatiotemporal variation in incidence and thus transmission risk.

For interventions that have already been tested and deemed effective, modeling can be used to determine optimal population coverage levels or to evaluate targeting of subsets of the population that exhibit distinct transmission dynamics and/or disease outcomes.

2. Evaluation of endgame strategies

Several tropical diseases are slated for local elimination or global eradication via mass administration of inexpensive drugs or via heightened awareness and the adoption of preventive behaviors. Endgame strategies to achieve and maintain elimination of disease in areas with low incidence can be evaluated using mathematical modeling. It is important to identify characteristics of remaining cases and incorporate them into the model through stratification of compartments or of individuals, in the case of an agent-based model, by risk of infection. Similarly, any individuals who are responsible for the bulk of transmission, such as due to genetic predisposition to high microbial loads, should be included within the model structure as superspreaders (Box 14.2). Modeling was used to inform onchocerciasis elimination efforts in West Africa. Administration of ivermectin was considered across transmission settings to determine what duration of treatment would be necessary to achieve local fadeout of disease [42].

(continued)

Box 14.1 (continued)

3. Guidance for health economic policy

Mathematical modeling can be used to guide decision-makers in developing policies. Cost-effectiveness analyses consider whether a particular intervention may not only have health benefits but also economic benefit. For instance, a preventative strategy may avert both adverse health outcomes as well as significant expenses associated with treatment. Traditional cost-effectiveness analysis does not incorporate ongoing transmission. Consequently, it has been assumed that when an individual is vaccinated, only that individual is protected. By combining cost-effectiveness analysis with a dynamic transmission model, the positive externality that vaccination has in terms of reducing transmission within the population can be incorporated. Such models have spurred change, such as in the case of the decision by the UK National Health Services to offer rotavirus vaccination for infants nationally [43], a policy previously deemed to be insufficiently cost-effective for the UK. As another example, results from a multi-host model of rabies transmission revealed that rabies vaccination of dogs is a cost-effective approach for preventing human disease in Tanzania, lending support for canine vaccination campaigns in the country [37].

4. Phylodynamics

Phylodynamic analysis of disease uses dated molecular sequencing of organisms to understand the implications of their evolutionary relationships on the epidemiology of disease. Specifically, such phylogenetic information provides insight into branching events or the timing of between-host transmission, such as zoonoses, and thus can inform parameters in models of disease systems [17]. Novel applications of phylodynamic approaches can estimate previously inestimable quantities such as the rates of underreporting or the numbers of subclinical infections by reconstructing transmission trees [44].

5. Big data applications

Big data refers to massive volumes of often unstructured data that require complex tools for their analysis. Evolving trends in everyday use of technology (e.g., social media, Google searches, cell phone usage) and improved computational capacity have provided novel sources of information about human interaction and behavior as well as ways to manipulate it, respectively. Open-source repositories have likewise made the data accessible for researchers of varied fields [17]. For instance, a recent modeling study investigating the global spread of *Zika virus* [45] incorporated large-scale flight itinerary data from the International Air Transport Association and a global population dataset to determine likely international dispersion pathways of the disease.

6. Probabilistic sensitivity analysis and uncertainty analysis

Uncertainty analysis (Box 14.2) and probabilistic sensitivity analysis (Box 14.2) have advanced the robustness of model fitting and the evaluation of model findings, respectively. Historically, model parameters have been calibrated to available data in an ad hoc manner [17]. Uncertainty analysis can be performed by Bayesian and semi-Bayesian approaches that evaluate a probable distribution of candidate parameter values from a prior distribution and/or their

(continued)

Box 14.1 (continued)

likelihood of producing model output that matches empirical data. Bayesian fitting methods [22] include Bayesian melding, Markov chain Monte Carlo integration, approximate Bayesian computation, particle filtering, and application of weighted Bayesian information criteria; Bayesian methods frequently supply a natural means to blend the empirical data being used for fitting (e.g., prevalence or incidence data) with other data from the literature (e.g., duration of infection, incubation period, and measures of relative infectiousness such as viral load). The posterior distributions of probable estimates that result from such methods characterize ranges of output that are appropriately influenced by but not entirely dictated by small sets of field data. Extension of Bayesian analyses to implement probabilistic sensitivity analysis is also straightforward and provides rigorous evaluation of model assumptions by considering the change in model output when drawing from ranges of empirically supported inputs [11]. These procedures of uncertainty analysis and probabilistic sensitivity analysis allow modelers to better guide public health decision-makers on the full range of potential transmission and intervention effectiveness outcomes.

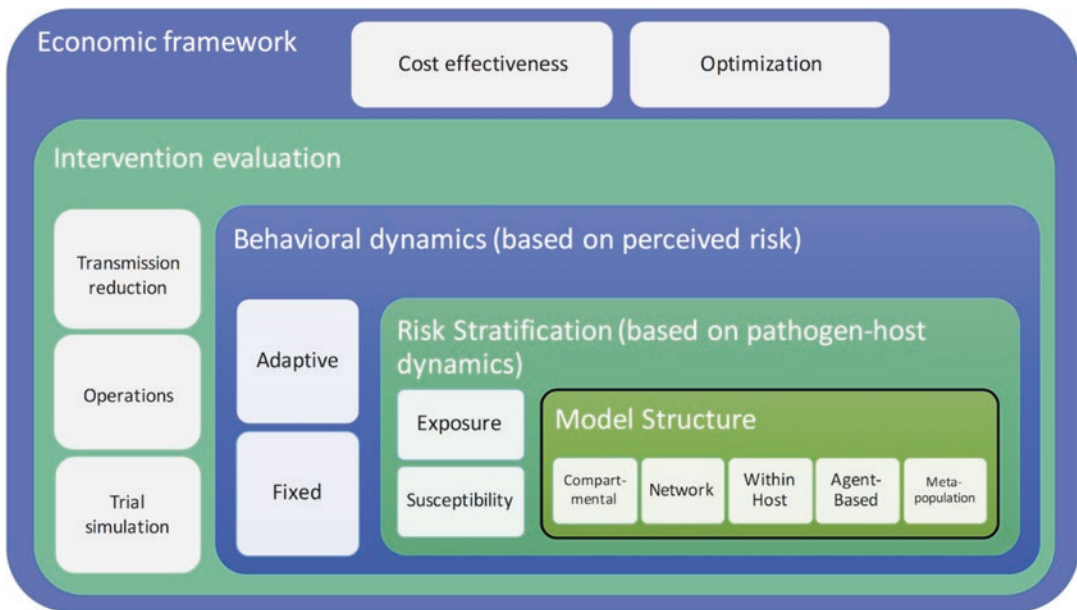


Fig. 14.1 Extension of transmission models to increase integration with field studies and public health policy relevance. The selected infectious disease model structure varies with the system being modeled and the research question, as outlined previously [9, 10]. A basic model design can be extended to include heterogeneity in biological susceptibility or exposure to infection. Exposure that varies due to behavioral dynamics can be accounted for as a fixed measure whereby a proportion of the affected population is involved in intervention with some level of adherence or as a dynamic measure that changes with time in response to increased awareness of disease. These extensions enhance the ability of the model to accurately estimate the effectiveness of interventions and thus estimate the costs of implementing strategies relative to changes in outcome measures

Box 14.2 Glossary of Terms

Superspreaders: Superspreaders are hosts that are responsible for the majority of transmission due to predisposition to high-intensity infection or behavior that results in heightened risk of exposure.

Network theory: In the context of infectious disease modeling, network models account for connections between social entities, such as individuals or groups of individuals who interact directly within the same household or community. Incorporating networks is one strategy that modelers can use to address systems where there is differential transmission risk for different groups within the network, where transmission occurs through close contact, or wherever the assumption of homogeneous mixing is inadequate.

Agent-based model: Agent-based models explicitly represent each individual in a system. For example, an outbreak of a nosocomial infection could be modeled by considering each patient and health-care worker within a hospital. The exposure history of each individual is tracked to generate probabilities of infection rather than modeling population-level rates of infection. However, the data requirements of informing parameters for each individual are often prohibitive.

Metapopulation model: A metapopulation includes many distinct subpopulations—often corresponding to spatial regions—that interact within and between themselves. Spatial metapopulation models can be used to represent the geographical dissemination of diseases, potentially parameterized by mobility data. For example, disease may be eliminated in one region, but due to migration from a neighboring region, it may later resurge.

Force of infection: The force of infection (FOI) is a measure of transmission. It is the rate at which individuals (e.g., human hosts, nonhuman hosts, vectors, body cells) become infected with a pathogen. In a mathematical model, the FOI is a mathematical function that determines the rate of transition between states. It is calculated using several parameters that account for the rate of contact between infectious individuals and susceptible (i.e., uninfected and nonimmune) individuals, the probability of successful pathogen transmission given contact, and the density or frequency of susceptible and infectious individuals in the population.

Behavioral economics: The field of behavioral economics focuses on how individual decisions are affected by psychological, social, and cognitive factors. Human behavior often deviates from what simple economic models predict rational decisions should be. These deviations can impact epidemiologically pertinent factors such as vaccination decisions or risk-related behaviors.

Prospect theory: Prospect theory is a behavioral economic theory of decision-making among probabilistic alternatives when the probabilities of different outcomes are known.

Likelihood fitting procedures: Likelihood fitting procedures involve developing an equation, or likelihood function, that calculates the probability of the data given a set of model parameters. Thus, parameters—whose values are unknown—can be estimated by maximizing the likelihood of observed data. For some multiparameter models with complex likelihood surfaces, identifying the set of parameters that maximizes the likelihood can be challenging. Diverse methods, such as the Newton-Raphson optimization algorithm, Markov chain Monte Carlo (MCMC), and simulated annealing approaches, have been developed to identify peaks of the likelihood surface corresponding to parameter values that are a good fit to data.

Probabilistic sensitivity analysis: Model outputs can be determined using best-fit point estimates for model parameters, as defined by experimental or field data or by fitting procedures. In a sensitivity analysis, the model outputs are then re-evaluated by varying parameter values

(continued)

Box 14.2 (continued)

slightly above and below the point estimate. The model is considered more sensitive to parameters for which small perturbations lead to large changes in model outputs. If individual parameter values are sampled in proportion to their probability, the sensitivity analysis is probabilistic.

Joint uncertainty distribution: Data rarely provide a precise value for a model parameter; that is, its value is considered uncertain. To best represent this uncertainty, a distribution of possible parameter values can be specified based on the data that are available. Model outcomes can be calculated based on serial draws from this distribution to determine a distribution of possible outcomes. If multiple model parameters are uncertain but correlated, then a joint distribution that simultaneously represents the effect of their combined uncertainty on the model output can be used to yield a distribution of outcomes that reflects uncertainty in parameterization.

Uncertainty analysis: An uncertainty analysis produces model output that takes into account uncertain parameter estimates for which available field or experimental data are missing or sparse. By iteratively drawing from a distribution of possible parameter values and running the model each time, a distribution of probable model output (versus just a single-point estimate) is produced. A full uncertainty analysis involves drawing from distributions for all uncertain parameters with each iteration of the model. When multiple outcomes are possible, a full uncertainty analysis can provide policy-makers quantification of which outcomes are more likely than others.

Game theoretic analysis: Game theory formalizes strategic decision-making in a group. The payoff of one's decision-making process depends on what other individuals decide, the information they have, the options they are presented with, and the perceived outcomes of each decision.

Nash equilibrium: In game theory, the Nash equilibrium refers to a solution to a problem where none of the decision-makers would benefit from opting for a different choice after becoming aware of other players' choices.

One common model structure is the compartmental model (Fig. 14.2), which divides a population into compartments depending on disease or demographic states. For instance, an SIR compartmental model categorizes individuals as susceptible (S), infected (I), and recovered (R). Individuals transition between compartments at rates that reflect population-level averages and according to difference equations (for discrete time) or differential equations (for continuous time). The SIR compartments provide a dynamic framework enabling current knowledge of historical disease prevalence and/or incidence, infection duration, and immunity to inform future predictions. They can be expanded on to arbitrary complexity—for instance, to model disease progression within the compartmental framework, the infection compartments can be divided into different phases. Depending on the disease, the stratifications may include a latency period modeled by moving susceptible individuals to a transitory compartment for those who have been exposed but are not yet infectious. In this way, the standard SIR model becomes an SEIR model (Fig. 14.2). For HIV, stratifying the infectious compartment I by thresholds of CD4 levels that interplay with viral load, symptoms, and transmissibility has often been incorporated into assessments of criteria for treatment.

14.4.2 Modeling Assumptions and Approaches

A basic assumption underlying compartmental models is that the hosts within a population mix homogeneously, such that individuals interact uniformly across compartments. Nonetheless, heterogeneity within host populations can be incorporated by stratifying compartments according

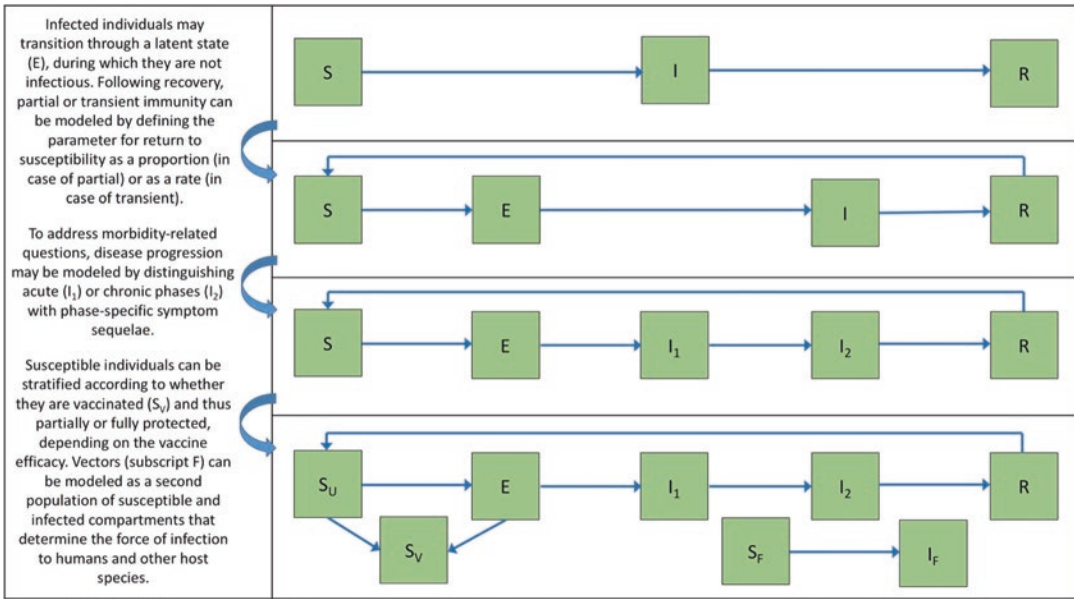


Fig. 14.2 Models are constructed with close accordance to the hypotheses that the models will be used to investigate. The structure of most disease transmission models is based on compartmentalizing individuals into categories of susceptible (S), infected (I), and—depending on the disease mechanism—recovered (R) with immunity. Additional model complexity could involve including a latency compartment (E) during which individuals are not infectious. Each arrow between compartments represents a directional rate of transition (or parameter) which is ideally informed by empirical data. Simple compartmental models are limited in their ability to capture complex disease systems; model extensions such as risk stratification and other model structures with finer units of transmission may be considered depending on the availability of sufficient data for parameterization

to, for example, age and risk factors that affect likelihoods of disease states. When a pathogen is composed of multiple antigenic strains that each elicit distinct immunological responses—such as is the case for pneumococci or human papilloma viruses—it could be important to stratify the immune state of the host according to strain-specific immunity. Stratifying by strain-specific immunity, for example, would permit accurate evaluation of the effectiveness of a vaccine that only covers a subset of the circulating strains. However, stratification should be used sparingly in deterministic compartment models, to reflect salient features of the epidemiological dynamics: as stratifications expand, the number of equations needed to describe transitions between compartments can exceed the number of host individuals in the population. If multifactorial stratification is necessary, it can become more efficient to adopt an agent-based framework (Box 14.2) in which the progression of each individual host is tracked through model states. In an agent-based model, each rate of disease transmission and progression is attributed specifically for every individual. However, agent-based models can only perform better to the extent that they can capture real information about contact mixing in the population. Availability of sufficient data with which to parameterize such individual-level specificities is a principal challenge in the development of agent-based models (Box 14.2).

For both compartmental and agent-based models, the “unit” represented could be a single human or nonhuman host or families or communities thereof, vectors that transmit pathogens between hosts, or body cells (as could be the case for within-host modeling). Other frameworks, including network models (Box 14.2), incorporate relationships between groupings of individuals, such as social connections or ecological niches of relevance to disease transmission. Scaling

up from representing a single population to a group of populations, metapopulation models (Box 14.2) can be used to incorporate spatial heterogeneity in disease risk factors. Within-host models can be used to evaluate immunological responses and pharmacokinetics of treatment by simulating dynamics at the cellular or even molecular level such that population-level trends can be inferred.

Models are sometimes criticized because they make such deterministic predictions in the face of many stochastic factors that are known to impact spread, such as the location of any initial zoonotic transmission, the behavior of initial cases in an outbreak, weather conditions, or relevant failures of public health infrastructure. Such stochastic factors have a pronounced impact on population dynamics when the number of infected is small, such as occurs in phases of disease emergence or extinction. Because the outcome of simulating deterministic models is always the same for a given set of parameters and initial conditions (e.g., population size and relative numbers of susceptible and infected individuals), incorporation of stochasticity into models enables probabilistic predictions that can quantify unlikely but highly important outcomes (e.g., persistence of disease despite a high chance of elimination by temporarily imposed measures). In stochastic simulations, individual transitions among states are counted explicitly and are probabilistically determined by random draws.

14.4.3 Model Outputs: Measure of Disease Transmission

A historically dominant measure of disease transmission is the basic reproductive number, R_0 , defined as the average number of secondary infections caused by a single primary infection in a naïve host population. If R_0 is greater than 1, an epidemic will typically invade a population. Sustained transmission is expected in the absence of control interventions and until the population develops sufficient herd immunity; even if the population develops herd immunity from infection, recurrent epidemics are likely to occur with the waning of immunity or with incoming births that are immunologically naïve. R_0 is a property of the causative agent that cannot be measured independent of the context in which contact between susceptible and infectious individuals occurs, reflecting the probability of pathogen transmission given contact as well as the duration of infectiousness.

While R_0 provides a threshold for disease emergence, it does not incorporate the length of time over which transmission occurs. For example, the R_0 values for influenza and for HIV have both been estimated at around 2 for specific settings [9], yet the epidemiological trajectories of these two diseases are dramatically different because the time period over which each infected host exposes two other hosts differs dramatically between the pathogens. Individuals with influenza remain infectious for about a week, while those with HIV can remain infectious for several months to years in the absence of treatment. Thus, to project epidemic curves and to estimate the timescales of outbreaks, the generation time (Box 14.2) of an infection is essential. Generation time, in turn, is determined by the within-host progression of infection as well as mode of transmission. R_0 can be elaborated on by quantifying the effective reproductive number, R_e , which is the average number of secondary infections caused by a single primary infection in a host population with some pre-existing immunity caused by vaccination or previous infection. Because pre-existing immunity is typically present to some degree, interventions that suppress R_e below 1 are predicted to control and ultimately eliminate the disease. This deterministic threshold is one basis for policy decisions regarding the level of interventions that should be implemented to eradicate a microbial disease. However, an intervention that brings R_e to a level just below 1 is only about 50% likely to actually do so [11] as its calculation is dependent on the accuracy of all parameters estimated.

Setting the level of intervention and calculating R_e for a set of samples from the joint uncertainty distribution (Box 14.2) of all the parameters provide the probability of disease elimination in that control scenario. Public health decision-makers—who realize that with imperfect data there is uncertainty associated with all policies—can then weigh the cost of intervention against the potential for failure based on the probabilities of specific outcomes occurring under stated scenarios of intervention [11].

Mathematical models have been developed to assess the epidemiological and health economic impact of alternative intervention approaches across a range of transmission scenarios and implementation strategies [12]. They are particularly valuable when used prior to investment in large-scale field trials or policy changes—that is, when they can be applied to evaluate likely reductions in disease burden or implementation costs. Additional innovations in modeling applications and methodologies (Box 14.1) have enabled the optimization of trial designs to maximize statistical power and feasibility [13], the evaluation of synergistic or antagonistic roles of coinfection in terms of morbidity risk [14], the incorporation of population behavior and perceptions toward vaccines and other interventions [3, 4], and the inference of previously immeasurable quantities such as the relative contributions of different host groups and subgroups to transmission [15, 16]. Compartmental, stratified compartmental, network, and agent-based models lie along a continuum of increasing resolution into individuality. Fine-grained models have an appeal because of the potential to account for inhomogeneities that cannot be easily captured with models that group individuals into classes. However, the more idiosyncratic, complex, or realistic we make a model, the less likely we will have the data to verify our assumptions are valid and the more computationally burdensome our analysis will be. Fortunately, a number of advances in infectious disease transmission modeling have broadened the utility of models by capitalizing on current knowledge and extant data to answer scientific questions and address infection control challenges.

14.4.4 Data-Driven Model Parameterization

The growing repositories of experimental, epidemiological, and clinical data, along with new and less traditional sources such as from social media, provide a wealth of information to enhance the capacity of disease transmission models to address public health questions [17, 18]. Such models are driven by a set of basic functions and parameters, including the rate of transmission, latency duration, infectious period, and waning of immunity. Values to inform parameters are derived from available surveillance data and published findings of observational studies, laboratory studies, and clinical trials, sometimes by using assumptions regarding applicability. For example, a natural history study on pneumococcal infections in children [19, 20] could be used to parameterize the attack rates for specific strains of *Streptococcus pneumoniae* across other age groups. Even small datasets that on their own do little to constrain a parameter can be informative using likelihood fitting procedures. See Box 14.2 that can draw upon that information in combination with other parameter constraints and primary epidemiological data such as prevalence and incidence rates. In Bayesian, or semi-Bayesian approaches, prior information based on small datasets can be usefully incorporated as a weak prior probability distribution for the parameter. Information from previous studies and the fit to current empirical data are increasingly being evaluated simultaneously with Bayesian or semi-Bayesian methods (see Box 14.1; [21, 22]). Incorporation of prior data that is informative about model parameters aids in the fitting of complex models with many parameters. Recent computational advances have rendered more complex models increasingly tractable; however, a healthy skepticism with regard to whether data is reliably informative must be applied to ensure predictions are well-founded. Moreover, with complexity come additional challenges adequately exploring

parameter space in the performance of model fitting. A complex model without sufficient data for parameterization can yield wildly inaccurate predictions—even when it has been constructed with a highly realistic, detailed structure and an excellent perceived fit to primary epidemiological data. Therefore, structural complexity and the number of model parameters should be kept to an informative minimum [23, 24].

Infectious disease datasets are frequently limited, in which case probabilistic sensitivity analysis (Box 14.2) of parameters and full uncertainty analysis (Box 14.2) of key outcomes can be conducted to provide an overview of the robustness of model findings. Sensitivity analysis can be performed by varying the values of parameters above and below the best point estimates and re-evaluating the model output. For instance, a cost-effectiveness study of rotavirus vaccination in Canada included a sensitivity analysis to evaluate how the net benefit of vaccination was affected by changes in population-level risk of rotavirus gastroenteritis [25]. Significant seasonal variation has been observed in rotavirus disease risk, and, even when accounting for variation in season of birth, rotavirus immunization remained a highly cost-effective strategy for British Columbia, Canada [25]. When results vary significantly with perturbation to an empirically uncertain parameter (i.e., a parameter whose value has only been established by minimal field or experimental data), a distribution of probable values—rather than just a single-point estimate—can be supplied to the model to generate a credible range of model outputs. In other words, point estimates alone do not quantify the extent of uncertainty in the estimate, and arbitrarily selected degrees of perturbation can poorly represent the effects of erroneous parameter values. There can, for instance, be substantial asymmetries in the uncertainty above and below-point parameter estimates. Quantifying the extent and asymmetry of uncertainty is important for determining the robustness of estimated effectiveness of control interventions that are fundamental to policy decision-making. For example, a study investigating pre- and posttreatment efficacy of mebendazole against soil-transmitted helminth infection in Tanzania reported the mean egg reduction rate (ERR) observed with the 95% confidence interval and sample size [26]. For these data, a point estimate approach would use the mean ERR as a measure of treatment efficacy, whereas an uncertainty analysis would involve iteratively simulating the model and randomly drawing an efficacy estimate from the full distribution of empirically observed ERRs. In the latter case, the predicted impact of a treatment campaign, such as changes in population-level infection intensity, would be represented as a range of all results calculated across the distribution of efficacy. When the effects of changes to the value of a parameter on outcomes are nonlinear within the range of uncertainty—as they often are in infectious disease models—the probabilistically weighted average outcome can be surprisingly different from the outcome calculated from the best point estimates. It should further be noted that, in addition to uncertainty in parameter values due to imperfect data, some parameters have inherent variability. Susceptibility to infection or treatment efficacy may vary due to factors such as genetic constitution or acquired immunodeficiency, such that drawing from a distribution of observed efficacy values in a compartmental model will reflect inherent heterogeneity in the parameter and thus in the system being modeled.

Along with the other methodologies presented in Box 14.1, the practices of probabilistic sensitivity analysis of individual parameters (see Box 14.2) and full uncertainty analysis of outcomes (see Box 14.2) are at the forefront of disease transmission modeling. Although the procedures are technical and computationally intensive, the benefit of these novel approaches is substantial to improve the robustness of findings for policy-focused research questions (Fig. 14.3). Because these analyses provide nuance to oracular-seeming assertions and therefore enhance the credibility of results, such recent methodological innovations have facilitated the wider adoption of transmission modeling as a component of analyses historically conducted within other disciplines, fostering transdisciplinary approaches to epidemiology and the development of public health policies.

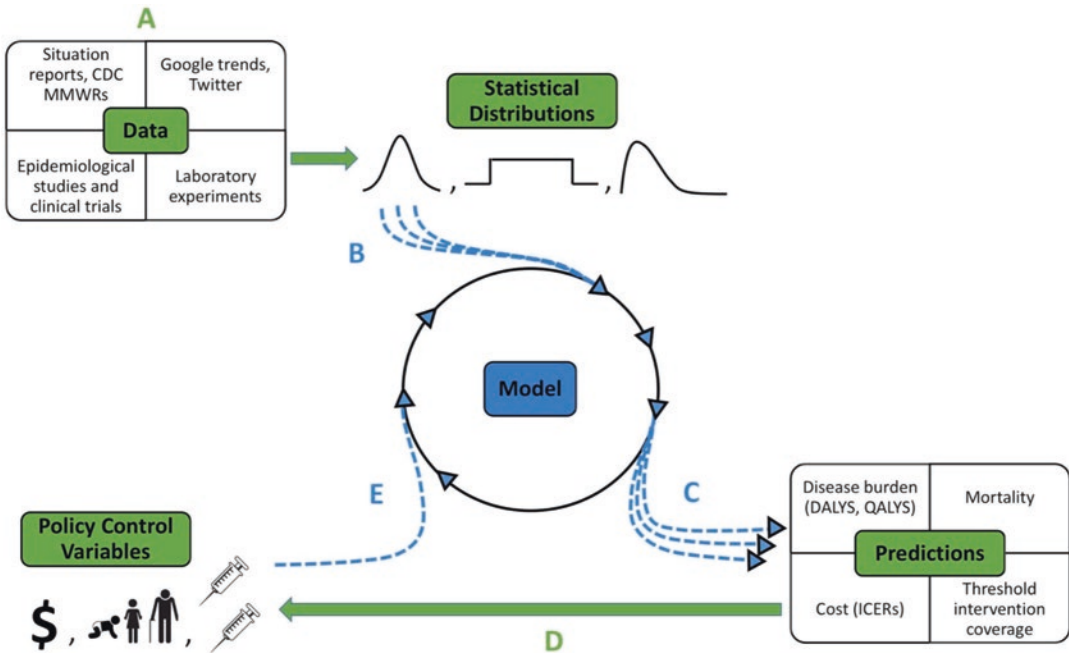


Fig. 14.3 Statistical inference to generate predictions about epidemiological trajectories, outcome measures, and likelihood of health policy effectiveness. Models can be better parameterized when (A) data drawn from a variety of sources, including both surveillance reports and published results of epidemiological studies, are used to inform model parameters. Rather than extracting single-point estimates, modelers can make full use of data by (B) constructing data-driven distributions of possible values for parameters and then sampling values from those distributions during iterative model runs. Iterations of the model are executed drawing probabilistically from distributions of data to generate (C) probabilistic distributions of predictions. Examples of the types of predictions that can be generated include estimates of disease burden and cost of intervention to inform policy decisions. Policy questions, such as whether to recommend a new vaccine, determine (D) specific variables to evaluate (e.g., cost-effectiveness). In a probabilistic framework, decision-makers are conveyed not only appropriate estimates of outcomes but also a sense of how much confidence they can have in those estimates. Settings for (E) variables that are under the presumed control of policy-makers can be pre-defined and can be further informed by initial model findings that lead to updated iterations of the model. Such policy variables could include cost, dosage schedule, and age of administration

14.4.5 Transdisciplinary Nature of Infectious Disease Modeling

Mathematical modeling can most effectively inform resource allocation and other health policy decisions when it is integrated with field epidemiology and basic science. In fact, the model-guided field-work paradigm [27] suggests that mathematical models can be incorporated into all phases of field research to improve our understanding of infectious disease ecology and complex biological systems. The iterative testing of hypotheses through empirical research and mathematical modeling while concurrently informing study design and model evaluation is a powerful approach for the study and control of disease (Fig. 14.3). The need for sequential data-gathering efforts and analysis and the consequent increased collaboration among experimentalists, policy-makers, and modelers are surmountable barriers to the integration of modeling evidence and public health policy. This iterative feedback approach has been increasingly advocated by researchers [12, 18] and has been demonstrated with applications in the evaluation of the rapid diagnostic test Xpert for tuberculosis [28]. An early modeling

study suggested that significant reductions in TB morbidity and mortality would be associated with increases in rates of TB diagnosis due to Xpert use and the assumed corresponding increases in treatment seeking. However, clinical trials in TB-endemic areas demonstrated lower-than-expected impact of the new diagnostic since treatment coverage in such areas was already very high [29]. The new data on TB treatment practices resulted in iterative improvements in the model so that it could more accurately evaluate new diagnostic technologies across different endemic settings [28].

As host behavior is increasingly incorporated into models of infectious disease and as host behavior becomes more important to infection control in an age of patient-inclusive decision-making, the social sciences—particularly psychology—also interface with epidemiology [4]. At this interface, survey data regarding perceived risks and benefits of vaccination as well as economic data on production and administration has been integrated with epidemiological modeling and game theoretic analysis (Box 14.2) to assess the extent that vaccination recommendations will be adopted by the public. For example, this multifaceted approach yielded predictions that “Nash equilibrium” (Box 14.2) vaccination levels would be much lower than the coverage that would be socially optimal [30], projections that were consistent with the national HPV vaccination coverage actually achieved among adolescent girls several years later [31].

Rather than considering a limited number of preconceived strategies, the application of optimization algorithms to epidemiological models enables the identification of the most favorable strategies from a virtually exhaustive range of feasible options. Despite computational challenges of updating model structure to assess all strategies for a given disease system, a modeling approach for the evaluation of options is often significantly more feasible than conducting diverse field trials. To inform health policy decisions, the criteria of optimization can incorporate economic and logistical considerations, as well as mortality and morbidity outcomes. For example, a model evaluating the optimal vaccine distribution to avert influenza outbreaks in the USA determined that the distribution of 63 million vaccine doses could reduce R_e and extinguish an outbreak that had similar transmissibility and mortality patterns as historical influenza pandemics [32]. Analyses of the age-structured model optimized vaccine allocation across a range of available doses and demonstrated that vaccination of school-aged children 5–19 years would be most effective by every outcome measure considered, infections, hospitalizations, mortality, years of life lost, and economic impact, a finding that was robust to parameter uncertainty [33]. This finding challenged the efficiency of the then-current strategy behind the estimated 85 million annual vaccine doses that were being administered for seasonal influenza. The results of this modeling study spurred a shift in influenza vaccination policy in the USA [34] to focus on children and teenagers, who are responsible for the most transmission, in addition to older adults, who are at greatest risk of influenza-associated morbidity and mortality.

Data-driven real-time applications in disease forecasting and surveillance are increasingly being incorporated into modeling [35]. Such applications include web-based services to simulate the spread of disease. They depend on digital disease surveillance, which involves Internet-based data from social media, news, health department, and other sources, to provide real-time information on outbreak detection and thus increase the timeliness and relevance of studies on disease trends and intervention strategies [36]. The Texas Pandemic Flu Toolkit (<http://flu.tacc.utexas.edu/>), for instance, has been used to forecast trends in pandemic flu throughout Texas using data from the state health department in near real time. The web-based dashboard allows for users to input different interventions, including antiviral treatment, vaccines, and public service announcements, and observe their impact on an unfolding epidemic.

Integration of economic analyses into these real-time applications and other epidemiological models provides policy-makers with evaluations of not only intervention efficacy but also economic impact and affordability. The cost-effectiveness of distributing Texas’ national stockpile of antivirals

during a flu pandemic could be determined based on forecasts of final size from the Texas Pandemic Flu Toolkit. Likewise, in close collaboration with field epidemiologists who understand implementation feasibility and real-world costs, modeling has been used to investigate the cost-effectiveness of “One Health” strategies, such as vaccinating dogs to protect against human rabies infection [37] and to design innovative interventions, including the treatment of schistosomiasis to reduce HIV susceptibility in sub-Saharan Africa [38]. These examples demonstrate the plasticity of modeling to consider novel intervention strategies that inform scientists of gaps in prevention and control measures and that address the changing demands on decision-makers.

14.5 The Future of Infectious Disease Epidemiology

With the emergence of zoonotic diseases, the spread of existing diseases to new settings, and global efforts toward elimination strategies for preventable infectious diseases, dynamic modeling of infectious disease is a powerful approach to address evolving questions and inform real-time decisions in emergency situations. Recent advances in microbial disease epidemiology have involved the application of molecular techniques for source tracking [39], the investigation into implications of human-driven changes in climate and landscape on the urbanization and emergence of disease [40], and the use of big data, such as from cell phone and social media usage, for exploring contact patterns and mobility trends in growing populations [41]. Combined, these approaches hold tremendous potential to improve our forecasting of disease through data collection and analyses. The instrumental role for quantitative epidemiology in these areas of innovation underscores the continued timeliness and significance of mathematical modeling as an application that informs and facilitates the translation of scientific findings into policy implementation.

References

1. GBD 2017 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392(10159):1859–922.
2. Reluga TC. Game theory of social distancing in response to an epidemic. *PLoS Comput Biol*. 2010;6(5):e1000793.
3. Manfredi P, D’Onofrio A. Modeling the interplay between human behavior and the spread of infectious diseases. New York: Springer Science & Business Media; 2013. 329 p.
4. Bauch CT, Galvani AP. Epidemiology. Social factors in epidemiology. *Science*. 2013;342(6154):47–9.
5. Oraby T, Thampi V, Bauch CT. The influence of social norms on the dynamics of vaccinating behaviour for paediatric infectious diseases. *Proc Biol Sci*. 2014;281(1780):20133172.
6. Wang Z, Andrews MA, Wu Z-X, Wang L, Bauch CT. Coupled disease-behavior dynamics on complex networks: a review. *Phys Life Rev*. 2015;15:1–29.
7. Sunstein CR, Zeckhauser R. Overreaction to fearsome risks. *Environ Resour Econ*. 2011;48(3):435–49.
8. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. *BMJ*. 2009;339:b2651.
9. Anderson RM, May RM. Infectious disease of humans. Dynamics and control, vol. 1. Oxford: Oxford University Press; 1991. p. 991.
10. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton: Princeton University Press; 2008.
11. Gilbert JA, Meyers LA, Galvani AP, Townsend JP. Probabilistic uncertainty analysis of epidemiological modeling to guide public health intervention policy. *Epidemics*. 2014;6:37–45.
12. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science*. 2015;347(6227):aaa4339.
13. Bellan SE, Pulliam JRC, Pearson CAB, Champredon D, Fox SJ, Skrip L, et al. Statistical power and validity of Ebola vaccine trials in Sierra Leone: a simulation study of trial design and analysis. *Lancet Infect Dis*. 2015;15(6):703–10.

14. NdeffoMbah ML, Skrip L, Greenhalgh S, Hotez P, Galvani AP. Impact of *Schistosoma mansoni* on malaria transmission in Sub-Saharan Africa. *PLoS Negl Trop Dis*. 2014;8(10):e3234.
15. Courtenay O, Carson C, Calvo-Bado L, Garcez LM, Quinnell RJ. Heterogeneities in *Leishmania infantum* infection: using skin parasite burdens to identify highly infectious dogs. *PLoS Negl Trop Dis*. 2014;8(1):e2583.
16. Miller E, Warburg A, Novikov I, Hailu A, Volf P, Seblova V, et al. Quantifying the contribution of hosts with different parasite concentrations to the transmission of visceral leishmaniasis in Ethiopia. *PLoS Negl Trop Dis*. 2014;8(10):e3288.
17. Woolhouse MEJ, Rambaut A, Kellam P. Lessons from Ebola: improving infectious disease surveillance to inform outbreak management. *Sci Transl Med*. 2015;7(307):307rv5.
18. Knight GM, Dharan NJ, Fox GJ, Stennis N, Zwerling A, Khurana R, et al. Bridging the gap between evidence and policy for infectious diseases: how models can aid public health decision-making. *Int J Infect Dis*. 2016;42:17–23.
19. Sleeman KL, Griffiths D, Shackley F, Diggle L, Gupta S, Maiden MC, et al. Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. *J Infect Dis*. 2006;194(5):682–8.
20. Smith T, Lehmann D, Montgomery J, Gratten M, Riley ID, Alpers MP. Acquisition and invasiveness of different serotypes of *Streptococcus pneumoniae* in young children. *Epidemiol Infect*. 1993;111(1):27–39.
21. Powers KA, Ghani AC, Miller WC, Hoffman IF, Pettifor AE, Kamanga G, et al. The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study. *Lancet*. 2011;378(9787):256–68.
22. Coelho FC, Codeço CT, Gomes MGM. A Bayesian framework for parameter estimation in dynamical models. *PLoS One*. 2011;6(5):e19616.
23. May RM. Uses and abuses of mathematics in biology. *Science*. 2004;303(5659):790–3.
24. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, et al. *Global sensitivity analysis: the primer*. Chichester: Wiley; 2008.
25. Fisman DN, Chan CH, Lowcock E, Naus M, Lee V. Effectiveness and cost-effectiveness of pediatric rotavirus vaccination in British Columbia: a model-based evaluation. *Vaccine*. 2012;30(52):7601–7.
26. Albonico M, Ramsan M, Wright V, Jape K, Haji HJ, Taylor M, et al. Soil-transmitted nematode infections and mebendazole treatment in Mafia Island schoolchildren. *Ann Trop Med Parasitol*. 2002;96(7):717–26.
27. Restif O, Hayman DTS, Pulliam JRC, Plowright RK, George DB, Luis AD, et al. Model-guided fieldwork: practical guidelines for multidisciplinary research on wildlife ecological and epidemiological dynamics. *Ecol Lett*. 2012;15(10):1083–94.
28. Menzies NA, Cohen T, Murray M, Salomon JA. Effect of empirical treatment on outcomes of clinical trials of diagnostic assays for tuberculosis. *Lancet Infect Dis*. 2015;15(1):16–7.
29. Theron G, Zijenah L, Chanda D, Clowes P, Rachow A, Lesosky M, et al. Feasibility, accuracy, and clinical effect of point-of-care Xpert MTB/RIF testing for tuberculosis in primary-care settings in Africa: a multicentre, randomised, controlled trial. *Lancet*. 2014;383(9915):424–35.
30. Basu S, Chapman GB, Galvani AP. Integrating epidemiology, psychology, and economics to achieve HPV vaccination targets. *Proc Natl Acad Sci U S A*. 2008;105(48):19018–23.
31. CDC – Teen Vaccination Coverage – NIS – Teen – Vaccines [Internet]. [cited 2016 Jan 4]. Available from: <http://www.cdc.gov/vaccines/who/teens/vaccination-coverage.html>.
32. Medlock J, Galvani AP. Optimizing influenza vaccine distribution. *Science*. 2009;325(5948):1705–8.
33. NdeffoMbah ML, Medlock J, Meyers LA, Galvani AP, Townsend JP. Optimal targeting of seasonal influenza vaccination toward younger ages is robust to parameter uncertainty. *Vaccine*. 2013;31(30):3079–89.
34. Children, the Flu, and the Flu Vaccine | Seasonal Influenza (Flu) | CDC [Internet]. [cited 2016 May 18]. Available from: <http://www.cdc.gov/flu/protect/children.htm>.
35. Brownstein JS, Freifeld CC, Chan EH, Keller M, Sonrick AL, Mearu SR, et al. Information technology and global surveillance of cases of 2009 H1N1 influenza. *N Engl J Med*. 2010;362(18):1731–5.
36. Salathé M, Freifeld CC, Mearu SR, Tomasulo AF, Brownstein JS. Influenza A (H7N9) and the importance of digital epidemiology. *N Engl J Med*. 2013;369(5):401–4.
37. Fitzpatrick MC, Hampson K, Cleaveland S, Mzimhiri I, Lankester F, Lembo T, et al. Cost-effectiveness of canine vaccination to prevent human rabies in rural Tanzania. *Ann Intern Med*. 2014;160(2):91–100.
38. NdeffoMbah ML, Kjetland EF, Atkins KE, Poolman EM, Orenstein EW, Meyers LA, et al. Cost-effectiveness of a community-based intervention for reducing the transmission of *Schistosoma haematobium* and HIV in Africa. *Proc Natl Acad Sci U S A*. 2013;110(19):7952–7.
39. Gire SK, Goba A, Andersen KG, Sealfon RSG, Park DJ, Kanneh L, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*. 2014;345(6202):1369–72.
40. Neiderud C-J. How urbanization affects the epidemiology of emerging infectious diseases. *Infect Ecol Epidemiol*. 2015;5:27060.
41. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. *PLoS Med*. 2013;10(4):e1001413.

42. Stolk WA, Walker M, Coffeng LE, Basáñez M-G, de Vlas SJ. Required duration of mass ivermectin treatment for onchocerciasis elimination in Africa: a comparative modelling analysis. *Parasit Vectors*. 2015;8:552.
43. Atkins KE, Shim E, Carroll S, Quilici S, Galvani AP. The cost-effectiveness of pentavalent rotavirus vaccination in England and Wales. *Vaccine*. 2012;30(48):6766–76.
44. Scarpino SV, Iamarino A, Wells C, Yamin D, Ndeffo-Mbah M, Wenzel NS, et al. Epidemiological and viral genomic sequence analysis of the 2014 ebola outbreak reveals clustered transmission. *Clin Infect Dis*. 2015;60(7):1079–82.
45. Bogoch II, Brady OJ, Kraemer MUG, German M, Creatore MI, Kulkarni MA, et al. Anticipating the international spread of Zika virus from Brazil. *Lancet*. 2016;387(10016):335–6.