SUPPLEMENT ARTICLE

Infectious Disease Modeling Methods as Tools for Informing Response to Novel Influenza Viruses of Unknown Pandemic Potential

Manoj Gambhir,[1,2,3,a] Catherine Bozio,[4,a] Justin J. O'Hagan,[2,3,a] Amra Uzicanin,[5] Lucinda E. Johnson,[6] Matthew Biggerstaff,[6] and David L. Swerdlow[7]

[1]Epidemiological Modelling Unit, Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia; [2]Modeling Unit, National Center for Immunization and Respiratory Diseases (NCIRD), Centers for Disease Control and Prevention (CDC), [3]IHRC Inc, [4]Graduate Program in Epidemiology and Molecules to Mankind, Laney Graduate School, Emory University, [5]Division of Global Migration and Quarantine, National Center for Emerging and Zoonotic Infectious Diseases, [6]Influenza Division, and [7]Modeling Unit and Office of the Director, NCIRD, CDC, Atlanta, Georgia

The rising importance of infectious disease modeling makes this an appropriate time for a guide for public health practitioners tasked with preparing for, and responding to, an influenza pandemic. We list several questions that public health practitioners commonly ask about pandemic influenza and match these with analytical methods, giving details on *when* during a pandemic the methods can be used, *how long* it might take to implement them, and *what data* are required. Although software to perform these tasks is available, care needs to be taken to understand: (1) the type of data needed, (2) the implementation of the methods, and (3) the interpretation of results in terms of model uncertainty and sensitivity. Public health leaders can use this article to evaluate the modeling literature, determine which methods can provide appropriate evidence for decision-making, and to help them request modeling work from in-house teams or academic groups.

The 2009 influenza A (H1N1) pandemic was one of the most closely tracked and studied epidemics in history. Traditional epidemiological methods, such as outbreak investigations and laboratory-based surveillance, were rapidly used to inform policy decisions [1–4]. These methods were enhanced by newer computational techniques such as bioinformatics and digital surveillance methods [5]. Simultaneously, substantial contributions to the literature were made in the area of infectious disease modeling (IDM) [6–10]. This article is a guide to the way in which (IDM) can contribute to policy discussions and decision-making in preparation for, or during, an influenza pandemic.

During an outbreak of influenza with pandemic potential, public health leaders ask a range of questions to inform situational awareness, help assess severity [11] and guide decisions that aim to control the spread and impact of disease. Critical questions include:

- What is the case-fatality ratio?
- What is the case-hospitalization ratio?
- When will disease incidence reach its peak?
- Who in the population should be prioritized for vaccination or antiviral treatment?
- How transmissible is the disease?
- What is the basic reproduction number (R_0)?

The accuracy with which these questions can be answered is time/data-dependent; as time passes and the outbreak progresses, more data become available to analyze. As such, a good compendium of methods for influenza outbreak analyses will be clear about the data requirements of each method and precisely when, during an evolving pandemic, the method might be most useful. Despite, and perhaps because of, the large accumulation of knowledge regarding the population

[a]These authors contributed equally to the data-gathering and writing of this manuscript.

Correspondence: Manoj Gambhir, PhD, Monash University, Melbourne, Australia (manoj.gambhir@monash.edu).

Clinical Infectious Diseases® 2015;60(S1):S11–9

DOI: 10.1093/cid/civ083

dynamics of influenza, it has become difficult for those in public health agencies to know how modeling methods may be employed during an outbreak. In particular:

- What questions can be answered by a specific modeling method?
- At what stage of a pandemic might a particular modeling method be used?
- What data are required by the modeling method?

In this article, we define IDM methods as techniques that include the mechanism of transmission of infection from an infected to an uninfected host. These are distinct from most methods in epidemiology because they explicitly model the transmission process and therefore the cause of infection; they therefore account for the "dependent" nature of infection, in other words, the fact that a major risk factor for infection is the population infection prevalence itself. IDM methods use data collected as part of routine surveillance (eg, clinical case counts, % polymerase chain reaction positives among tested clinically diagnosed cases, time of symptom onset), but they also often suggest new datasets that investigators might collect, such as detailed household-based questionnaires with an emphasis on determining exposure time windows [6, 7]. Furthermore, the mechanistic assumption of person-to-person transmission used in IDM methods is a strong one and can allow estimates to be made with less data than otherwise.

Here, we collate and describe some of the most useful and well-tested IDM methods for use prior to, or during, an influenza pandemic. Our intent is to facilitate communication between public health practitioners and modelers during different stages of the response to pandemic threats, by reviewing these methods and linking them to the timeline of a pandemic response.

METHODS

We searched the published literature to produce an inventory of methods used to extract actionable information from influenza surveillance or study data. Due to the size of the modeling literature, we confined our search to articles reporting for analyses of the 2009 influenza A (H1N1) pandemic, and we augmented this search with relevant articles outlining IDM methods, whether purely theoretical or previously used for seasonal influenza analysis. Only articles that had been peer-reviewed were included.

We used the search term "H1N1 & {modeling OR modelling} & pandemic" in PubMed to find articles that described IDM methods pertaining to the 2009 H1N1 influenza pandemic. More specifically, we selected articles that employed IDM methods to answer questions on the basic epidemiology of influenza, and the projected benefit of vaccination, antiviral use, and nonpharmaceutical interventions (eg, school closures). We read those articles where the abstracts indicated that the method and research question were relevant to pandemic modeling. The search was conducted in April 2014, and the articles retrieved from PubMed were imported into the bibliographic reference package, EndNote version X6.

Data were extracted from the selected articles and summarized in a table that was structured to be accessible to the public health community; we listed key questions of interest to public health leadership and specific IDM methods that could answer these questions. We categorized the questions into a framework comprised of 4 components of a pandemic influenza response: (1) the basic epidemiology of the novel disease, (2) impact of vaccination, (3) effect of antivirals, and (4) the role of community mitigation. For each of the IDM methods that can be used in response to the questions, we assessed *when* during an evolving pandemic the method might be used (ie, at what point might there be sufficient data). These time-points were categorized as: P—Pre-pandemic (ie, preparedness planning); E—early pandemic (prior to the peak of cases); L—late pandemic (after the peak in the incidence curve and including post-pandemic assessment of the effectiveness of interventions that were implemented).

A further important feature of the surveyed approaches was the mathematical style of the method, which we categorized as 'statistical' or 'simulation'. Statistical methods are those that can be used to estimate or calculate the values of epidemiological parameters from available data. Simulation methods, on the other hand, use available parameters, taken from the literature or expert opinion, to run epidemic simulations forward in time. Finally, we give an indication of the time requirement for model development and execution on a computer (usually a desktop, though some methods such as individual-based models are often run on computer clusters). Note that these requirements do not include the time needed for data collection and cleaning, and additional time is required for this. The availability of software packages specific to the analysis may reduce some of these estimates.

Very few of the reviewed articles contained precisely identical methods, and we generally listed only 1 or 2 example articles for each approach. For example, a study using a stochastic age-structured susceptible-exposed-infected-recovered (SEIR) model and another using a deterministic one were both categorized as " 'age-structured SEIR model.' " On the other hand, we felt that the large number of new methods that have been developed to calculate the reproduction number, either at the beginning of a pandemic (ie, the basic reproduction number, R_0) or in real-time for an ongoing pandemic (ie, the effective reproduction number, R_e), were sufficiently different from one another to be listed separately.

Table 1. Key Questions Related to Pandemic Preparedness and Response That Infectious Disease Modeling Methods Address

Questions	Pandemic Stage			Method			
	P	E	L	Sim	Stat	Analysis Time Commitment	Method
Epidemiology							
What is the basic reproduction number (R_0) and the current value, or the time course, of the effective reproduction number (*Re*)?		X	X		✓	<1 mo	Growth rate of case incidence curve
		X	X		✓	<1 mo	Infection tree reconstruction
		X	X		✓	<1 mo	Richards population growth model
		X			✓	<1 mo	Chain binomial model
		X	X		✓	<1 mo	Case renewal process
		X			✓	<1 mo	Influenza genetic sequence analysis
		X	X	✓		<1 mo	Age-structured SEIR model
		X	X		✓	<1 wk	Maximum likelihood estimation
		X		✓	✓	<1 mo	Coalescent analysis
		X		✓		<1 wk	Next generation matrix
What is the predicted peak number of cases and time? What is the predicted cumulative number of cases over the epidemic (ie, final attack rate)?		X		✓		<1 mo	Age-structured SEIR model
				✓	✓	<1 mo	Digital surveillance methods
What are the possible spatiotemporal patterns of spread of the infection?		X		✓		>>1 mo	Individual-based model
				✓		<1 mo	Metapopulation model
What was the likely sequence of spatiotemporal spread of infection since the outbreak began?		X			✓	<1 mo	Infection tree reconstruction & travel pattern modeling
What is the severity of the virus(es) (ie, case-hospitalization/death-rate) accounting for ascertainment biases (eg, more likely to detect severe cases)?		X	X	✓		>>1 mo	Individual-based model
					✓	<1 mo	Bayesian evidence synthesis
					✓	<1 wk	Incidence curve backcalculation
What is the transmission probability of the virus?		X		✓		<1 wk	Contact rate matrices & SEIR model
What is the serial interval of the disease?		X	X		✓	<1 wk	Serial interval estimation
		X	X		✓	<1 wk/<1 mo	Maximum likelihood estimation (R_0)
What was the incidence of infection in a recent period?		X	X		✓	<1 wk	Incidence curve backcalculation
				✓	✓	<1 mo	Digital surveillance methods
How transmissible is the virus?		X			✓	<1 mo	Branching process analysis
					✓	<1 mo	Maximum likelihood estimation (household)
Can the changing circulating subtypes be predicted?		X	X	✓		>1 mo	Phylodynamic model
					✓	<1 mo	Antigenic cartography
What is the duration of immunity to the virus?			X	✓	✓	>1 mo	Time Series SIR model
Antivirals							
What might be/is the impact of antiviral distribution?	X	X		✓		<1 mo	Age-structured SEIR model
				✓		<1 mo	Metapopulation model
				✓		>>1 mo	Individual-based model
How should antiviral drugs be prioritized for distribution among population subgroups to: (a) maximize direct protection for those uninfected; (b) minimize infectiousness for those infected; (c) prevent the development of resistance? (d) prevent severe outcomes and shorten duration of illness among those infected	X	X		✓		<1 mo	Age-structured SEIR model
				✓		<1 mo	Metapopulation model
						>>1 mo	Individual-based model
Vaccination							
What is the potential impact of vaccination?	X	X		✓		<1 mo	Age-structured SEIR model
				✓		<1 mo	Metapopulation model
				✓		>>1 mo	Individual-based model

Table 1 continued.

Questions	Pandemic Stage			Method			
	P	E	L	Sim	Stat	Analysis Time Commitment	Method
When should vaccination be introduced? What is the impact of vaccine timing on the benefits of vaccination?	X	X		✓		<1 mo	Age-structured SEIR model
				✓		<1 mo	Metapopulation model
				✓		>>1 mo	Individual-based model
Who in the population should be vaccinated when the number of doses is limited?	X	X		✓		<1 mo	Age-structured SEIR model
				✓		<1 mo	Metapopulation model
				✓		>>1 mo	Individual-based model
				✓		<1 mo	Network model
Can next season's subtype combination be predicted in advance of its appearance, for more efficient vaccine production?		X		✓		<1 mo	Phylodynamic model
				✓		<1 mo	Antigenic cartography
Can vaccine effectiveness be predicted?		X	X		✓	<1 wk	Antigenic distance
Community Mitigation							
What is the appropriate selection, timing, and duration of community social-distancing measures to achieve the desired outcome (eg, delay the peak of epidemic to buy the time for implementation of other interventions (eg, vaccination), and/or reduce peak prevalence)?	X	X		✓		<1 mo	Age-structured SEIR model
				✓		<1 mo	Metapopulation model
				✓		>>1 mo	Individual-based model
What was the effect of the mitigation measures that were implemented?			X		✓	<1 mo	Partially observed Markov model

The questions are grouped into 4 categories: Epidemiology, Antivirals, Vaccination, Community Mitigation. The questions are also loosely grouped into those that may be asked pre-pandemic (P), early pandemic (E), and late pandemic (L, including post-pandemic) and also by the type of mathematical method, namely, simulation method (Sim) and statistical/mathematical method (Stat) (an explanation of these terms is included in the main text). The table also includes an indication of the time commitment, which is the estimated time taken to conduct an analysis for a researcher experienced with the method after the data have been cleaned. In the final column, we provide the names of the modeling methods that can be used (these names are explained in Table 2). Abbreviations: SEIR, susceptible-exposed-infected-recovered; SIR, susceptible-infected-recovered.

To reduce the main "questions" table (Table 1) to a more manageable size, we created a second complementary table (Table 2) to provide a more detailed description of each of the IDMs, indicating the type of data that are needed, and 1–2 references from our literature review that outline the method.

RESULTS

We present our overview of policy-related questions and IDM methods in Table 1, which is structured to emphasize the questions that might be asked in preparation for, during, or after an influenza pandemic.

Policy makers who need to know epidemiological quantities, such as the value of the basic reproduction number, R_0, have a large choice of methods, many of which can be used with the same data set and can therefore be compared with one another. For example, a policy maker needing to know the transmissibility of a novel influenza virus would consult the relevant questions in Table 1 (e.g., the questions relating to the basic reproduction number and the transmission probability). These table entries show that there are 10 methods for the calculation of R_0 and 2 for the calculation of the transmission probability and that all but one of these methods can be used at the early stages of a pandemic.

As another example, for pandemic planning or decision making that could lead to the production of influenza vaccine, policy makers want to know what the effect of a vaccination program might be. Table 1 indicates that there are 3 IDM methods that could be used to answer this question. Each of these methods are simulation-based and may take a month or longer to implement (a potentially very important fact, especially if a pandemic is already underway). Looking up these 3 methods in Table 2 shows that their data-needs are also extensive, and so the time devoted to data-gathering should be added to that needed for model implementation.

DISCUSSION

In this article, we aim to bridge the gap between the needs of public health practitioners concerned with influenza pandemics and the rapidly growing number of epidemiologists

Table 2. Description of the Modeling Methods Listed in Table 1

Method	Description	Data Needed
Age-structured SEIR model	A compartmental model in which hosts are grouped into population compartments composed of their age-group and their infection status, eg, an SEIR model. These models can be deterministic or stochastic, and the transitions between infection states are governed by contact and recovery rates [10, 12].	Case incidence stratified by age, contact matrix by age, cross-sectional serosurveys, physician visit/hospitalization rates to calculate symptomatic proportion/disease reporting rate/proportion immune, severity of infection across risk groups, initial number of infected individuals (or date on which the first infected individual was introduced into the population).
Antigenic cartography	A method for quantifying and visualizing the antigenic evolution of the influenza virus according to antigenic distances [13].	Influenza virus genetic sequences, antigenic distances between subtypes (using eg, hemagluttination inhibition assay).
Antigenic distance	Antigenic distances of proposed vaccine strains from predicted dominant circulating strain(s) are correlated with prior years' vaccine effectiveness estimates [14].	Hemagluttination inhibition assay distances of potential circulating strain(s) and record of vaccine effectiveness from prior years with amino acid sequences of past vaccine strains and dominant seasonal strains
Bayesian evidence synthesis	Prior knowledge and distinct surveillance data sources are combined to estimate epidemiologic quantities (eg number infected, case-hospitalization rate) [9].	Repeated cross-sectional serosurveys, numbers and dates of onset of confirmed cases, symptomatic cases, hospitalizations, intensive care admissions, dates of severe outcomes.
Branching process analysis	Branching process theory is used to estimate the number of offspring of primary cases [15]. The generation time distribution between households and incidence of infection of households [16] is estimated.	Contact tracing data, surveillance datasets, R_0 population distribution (ie, the probability associated with an individual in the population generating R_0 secondary cases at the start of the epidemic).
Case renewal process	Initial cases are modeled as a renewal process, which is a generalization of the Poisson process in which the time between cases is random and arbitrary, but independent and identically distributed [8, 17].	Case incidence time series (infection/hospitalization/death).
Chain binomial model	Initial cases are modeled as a discrete time chain of infections from one individual to another with probability of infection, or escape from infection, calculated using the binomial probability distribution [18].	Case incidence time series (infection/hospitalization/death).
Coalescent analysis	A Bayesian phylogenetic "coalescent" model is fitted to genetic sequence data obtained from isolates sampled from the infected population [19]. Growth rates of the epidemic are used to calculate the reproduction number and the "time to the most recent common ancestor" approach is used for estimating the start time of the outbreak.	Influenza genetic sequences and sampling times.
Contact rate matrices & SEIR model	Contact rate matrices [20] are used in a stochastic or deterministic SEIR model [20, 21]; the transmission probability as a function of age is inferred by fitting to age-structured incidence data.	Infection incidence over age, contact matrix by age.
Digital surveillance methods	Internet, social media, and mobile phone data are used to track disease spread (eg, by analyzing data from internet search queries and social media sites that are pertinent to illness) [22].	Google search queries, Twitter data, city/region-wide data on contacts, climate data
Genetic sequence analysis	The diversity between sampled influenza genetic sequences is used to infer the size of past epidemic spread [23].	Influenza genetic sequences with known population sampling rate and sampling times.
Growth rate of case incidence curve	An exponential growth model is fitted to early case incidence curve to estimate the initial epidemic growth rate [24].	Case incidence time series (infection/hospitalization/death).
Incidence curve backcalculation	A calculation that uses the disease or infection prevalence curve (eg, the seroprevalence) and the disease natural history to reconstruct the most probable past time course of the incidence of disease [25].	Repeated cross-sectional serosurveys, hospitalization incidence from the same population.
Individual-based model	A model in which the interacting elements are distinct individuals (rather than frequencies or densities of individuals, as with compartmental models). Individual-based models can be enormously computationally intensive because they often include a very large number of individual units (eg, in US-wide influenza models with 300+ million people, their travel patterns and household structures) [26, 27].	Social/age-mixing patterns and contact frequencies at home, work, weekdays/weekends (ie, contact matrix by age [14]); household size distribution; population density; survey data (eg, on behavior change following the implementation of social-distancing measures).

Table 2 continued.

Method	Description	Data Needed
Infection tree reconstruction	In the presence of missing infection or clinical onset times, a likely infection tree may be reconstructed from an ensemble of several candidate trees. The more information that may be provided to this technique (eg, generation and incubation time distributions as well as, more recently, genetics data) the better it performs. The reproduction number can be directly read off an ensemble of trees [28–32].	Ideally, every infection case and the time of infection. Any subset/variation of this data may also be used (eg, every case symptom onset or infection time, hospitalized cases and time (and case hospitalization rate), death times (and case fatality rate)
Infection tree reconstruction and travel pattern modeling	The infection tree is reconstructed and the spatiotemporal spread of the virus is inferred [33]. Spatiotemporal inferences can be validated using passenger travel data.	Influenza genetic sequences, dates and location of sampling, passenger flight data.
Partially observed Markov model	A Markov model that consists of the underlying process that represents the phenomenon of interest (eg, infection in the population) as well as an observation process, representing observations of the underlying phenomenon (eg, the surveillance system) [34–36].	Case incidence time series (infection/ hospitalization/death).
Maximum likelihood estimation (R_0)	A method for simultaneously estimating the basic reproduction number and the serial interval during the exponential growth phase of the outbreak [37].	Case incidence time series (infection/ hospitalization/death).
Maximum likelihood estimation (household)	An estimate of the household secondary attack rate is made by maximizing the likelihood of a data set of secondary case onset times within households [35]. The basic reproduction number is estimated by simulating an outbreak with the same secondary attack rate.	Primary and secondary case onset data within households, dates of infection/symptom onset, zoonotic exposure data (if applicable)
Metapopulation model	A model in which the total population is made up of subpopulations, connected to each other by interpopulation migratory rates. Such a structure can be used to model geographically distinct populations and their interactions [38].	Contact matrix by age, generation time distribution, rates of migration between population patches.
Network model	A simulation method that incorporates data on the contact structure of the population in which infection is spreading [39].	Heterogeneous contact rate distribution.
Next generation matrix	A matrix describing the average number of new transmissions (epidemiological "births") and transitions between states during one transmission cycle [20]. Determining the largest eigenvalue of this matrix gives the basic reproduction number.	Proportion of population initially immune (eg, from early serosurvey or pre-existing samples), R_0, generation time, contact matrix.
Phylodynamic model	A model in which both evolutionary (ie, genetic) and epidemiological processes relevant to pathogen transmission are explicitly taken into account [40, 41]. The transmission model may be of SEIR form, whereas the genetic model may be, for example, a coalescent phylogenetic model.	Influenza genetic sequences, antigenic distances between subtypes (eg, hemagluttination inhibition assay).
Richards population growth model	The basic reproduction number R_0 is estimated by fitting infection curves to incidence data from early in outbreak [8] using the Richards growth model, which is a generalized logistic function of time.	Case incidence time series (infection/ hospitalization/death).
Serial interval estimation	Candidate probability distributions (eg, gamma or lognormal) are fitted to frequency histograms of observed serial intervals [42].	Infector-infectee pairs (eg, from contact tracing early in the pandemic or a household study later in the pandemic), case incidence time series (infection/ hospitalization/death with case hospitalization rate or case fatality rate as appropriate).
Time Series SIR	A state space SIR model [43] that initializes the number of susceptible individuals and fits the modeled incidence to incidence data by varying, for example, the transmission rate.	Case incidence time series (infection/ hospitalization/death).

These methods are explained briefly in column 2 along with 1–2 references in which the method is outlined in detail. Column 3 gives an indication of the kind of data that are needed by the method. Previously unexplained technical terms are described in the Overview of Terms (Box 1). Abbreviations: SEIR, susceptible-exposed-infected-recovered; SIR, susceptible-infected-recovered.

who use IDM methods to investigate such threats. More specifically, we have developed a guide to facilitate communication between public health practitioners and modelers. The guide describes some of the most useful and well-tested IDM methods for use prior to, or during, an influenza pandemic within an explicit framework based on the progression

Box 1. Overview of Terms

Antigenic cartography: A combination of a multidimensional clustering algorithm and a visualization to allow influenza hemagluttinin inhibition assay distances between influenza strains to be viewed as two-dimensional maps.

Backcalculation: Quantities of epidemiological interest, such as the incidence of infected cases, often can only be inferred through indirect evidence, in other words, the available surveillance data. For example, if these data are the numbers of influenza-associated hospitalizations, then the incidence of infection can be calculated by estimating the incidence—at a prior time point—of the infections that led to the recorded hospitalizations. This hospitalized number needs to be inflated by the case-hospitalization ratio to obtain the infected number. Models are often only defined unambiguously forward in time and so extra assumptions are often needed to perform the reverse time calculations.

Basic reproduction number: The average number of infections that a single infectious case gives rise to at the beginning of an epidemic.

Branching process: A mathematical process in which nodes at one time point (or generation) give rise to further nodes at the next generation. The probabilities of proliferation or diminishment of nodes with each generation are described by well-defined statistics and several useful mathematical results exist. The process of infection propagation may be framed in branching process theory, where nodes represent infected individuals and the resulting branching process graph can be visualized as an infection tree.

Compartmental model: A model in which the population of interest is divided, at a given point in time, into specific compartments. The most widely used of these is the S-E-I-R model in which the population occupies one of the following mutually exclusive states: S-Susceptible, E-Exposed, I-Infected, R-Recovered.

Deterministic model: A model in which there are no random processes; consequently, all model runs give identical results. Outcomes from such models may often be viewed roughly as averages of more realistic stochastic models.

Effective reproduction number: The average number of infections that an infectious case gives rise to during an epidemic. This number is equal to the basic reproduction number at the beginning of an epidemic.

Fitting to data: The procedure by which a model's parameters are altered so that one or more of its outputs grow closer to a designated data set. Methods used to fit a model to data include minimization of squared differences or modern statistical inferential methods such as maximization of likelihood, or Markov Chain Monte Carlo using Bayes Theorem.

Generation time: The average time that elapses between the infection of a primary case and a secondary case.

Initial conditions: The starting values of each of the state variables at the beginning of a simulation run (eg, the number susceptible, infected, recovered).

Initial growth rate: The rate of increase in the number of infections at the beginning of an epidemic.

Markov model: A model in which the state of the system at a particular point in time is dependent only upon the state of the system at the previous point in time.

Parameterization: The process by which values or distributions of values are obtained for model parameters. Methods to do this include statistical inference by fitting models to data, and literature searches for plausible parameter values, though most parameterizations involve both.

SEIR model: See "Compartmental model."

Serial interval: The average time that elapses between the clinical onset of a primary case and that of a secondary case.

Stochastic model: A model that contains random processes from various sources, such as demographic and observation/reporting. Since these models contain randomness, each run on a computer is unique, and an ensemble of runs is necessary to produce an average model trajectory and uncertainty range for each model output.

State Space Model: A discrete-time Markov form of a compartmental model, often simplified through linearization of the full nonlinear model, that allows simplified analysis, simulation and fitting to data.

Statistical inference: See "Fitting to data."

of an outbreak and the requirements of the pandemic response.

Our guide to IDM methods during a pandemic consists of 2 tables. Public health practitioners can use Table 1 to determine which IDM methods are applicable for the specific questions they have. They can then refer to Table 2, which provides descriptions and data requirements for these methods. In some cases readers may find that they first need to use a method to calculate a quantity that can then be fed into another method. For example, the reproduction number might be calculated using a time-series of disease cases, and this can then be used to build a simulation model in which interventions such as social-distancing might be tested.

To execute the methods described in our guide requires epidemiologists with significant modeling experience; these specialists need to understand the relevant literature, convert a published mathematical method into computer code, request data in the correct form for model calibration, and then interpret model outcomes with appropriate consideration given to uncertainty. While user-friendly software implementing IDM methods remains largely unavailable, a critical need exists for personnel in public health agencies to occupy translational roles, benefiting from training in mathematical modeling techniques as well as pandemic preparedness and response, to assist in the selection and deployment of appropriate IDM methods.

We found that a large number of new IDM methods are under development to answer questions relating to basic epidemiology, such as the calculation of the basic reproduction number. Quantities such as R_0 are indirectly related to the specific questions of policy makers; once estimated, they can be used to parameterize simulation models, which can then be used to

perform scenario analyses for a variety of public health interventions. One of the main reasons for the proliferation of methods for calculating epidemiological quantities is the tendency for authors to publish novel scientific methods rather than apply established methods to answer policy-specific questions. This highlights the translational gap between novel IDM methods and public health practice; efforts directed toward bridging this gap would be valuable to public health agencies, and this is the major purpose of the present article.

While the methods included here can be broadly divided into 2 mathematical categories—statistical and simulation—this categorization is not mutually exclusive. For example, an SEIR model simulates the spread of infection; however, SEIR model parameter values (eg, duration of infectiousness) are often estimated by running the model within a statistical framework, in which the model outputs are compared with corresponding data; the best-fitting parameter values are those which minimize the discrepancy between the model output (eg, an incidence curve) and the surveillance data (eg, weekly illness incidence). Individual-based models are almost exclusively simulations, because they are so computationally intensive that incorporating them into a statistical framework is prohibitively time-consuming (for exceptions see [44, 45]); also, the number of parameters individual-based models contain can be so large that it becomes infeasible to attribute changes in summary statistics (eg, number of cases averted by vaccination) to changes in specific parameter values (eg, contact rates between age groups).

There are very few published studies that attempt to collate the IDM literature with the aim of bringing together the *questions* that are asked by leaders during a public health emergency and the *methods* that can be used to address them. More often, reviews have presented the *results* of modeling studies in a digestible format (such as those for SARS [46] or HIV/AIDS [47]). Our survey of the landscape of IDM methods applicable to pandemic influenza (which may be the infectious disease area in which the field of IDM is advancing most rapidly) differs from previous reviews by placing at its center the questions that might be asked by public health leadership to improve situational awareness during, or prior to, an influenza pandemic. At present, traditional approaches to some of the questions we consider (eg, laboratory work, epidemiological studies) may at times be more appropriate than modeling methods, but modeling is a complementary activity and is not a substitute for other analyses.

The main limitation of our guide is that it will need to be updated as new techniques move from the research frontier to become better-established as usable tools. As such, this review represents a snapshot of currently available methods. Nevertheless, we hope it will be useful to both public health practitioners and IDM modelers. The time-dependence of analytical methods is an extremely important factor that we have addressed in this article but that we should reiterate here: cases of influenza disease naturally take time to accrue, and it is only once a sufficient number have been observed using reliable surveillance methods that we can make model-based inferences with any degree of confidence.

There is a significant communication gap between these 2 groups, public health practitioners and IDM modelers, which needs to be bridged in both directions. This can only be achieved if both parties make a concerted effort to understand the needs, strengths, and limits of one another. We hope articles such as this one serve as a possible catalyst to this dialogue. Because there have been several pandemic threats over the last decade from influenza alone, public health agencies will be better served by embracing new analytical methods, with a proper appreciation for their uncertainty and the data they require. IDM modelers can better serve such agencies by taking the time to ensure their methods are accessible and address the specific needs of public health practitioners, and not only their own research priorities.

Notes

Acknowledgments. M. G. was funded through a grant from the US National Institutes of Health, National Institute of General Medical Sciences (grant number 1U01GM110721-01). C. B. would like to thank Emory University's Molecules to Mankind program. All other authors were employed by the US Centers for Disease Control and Prevention (CDC) when this work was conducted.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the CDC.

Supplement sponsorship. This article appears as part of the supplement titled "CDC Modeling Efforts in Response to a Potential Public Health Emergency: Influenza A(H7N9) as an Example," sponsored by the CDC.

Potential conflicts of interest. All authors: No potential conflicts of interest.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Outbreak of 2009 pandemic influenza A (H1N1) at a school—Hawaii, May 2009. MMWR Morb Mortal Wkly Rep **2010**; 58:1440–4.
2. Lessler J, Reich NG, Cummings DA, Nair HP, Jordan HT, Thompson N. Outbreak of 2009 pandemic influenza A (H1N1) at a New York City school. N Engl J Med **2009**; 361:2628–36.
3. Iuliano AD, Reed C, Guh A, et al. Notes from the field: outbreak of 2009 pandemic influenza A (H1N1) virus at a large public university in Delaware, April–May 2009. Clin Infect Dis **2009**; 49:1811–20.
4. Han K, Zhu X, He F, et al. Lack of airborne transmission during outbreak of pandemic (H1N1) 2009 among tour group members, China, June 2009. Emerg Infect Dis **2009**; 15:1578–81.
5. Kass-Hout T. Biosurveillance. Hoboken: CRC Press, **2010**:1. online resource (382 p.).
6. Cauchemez S, Bhattarai A, Marchbanks TL, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proc Natl Acad Sci U S A **2011**; 108:2825–30.
7. Cauchemez S, Donnelly CA, Reed C, et al. Household transmission of 2009 pandemic influenza A (H1N1) virus in the United States. N Engl J Med **2009**; 361:2619–27.

8. Mostaco-Guidolin LC, Bowman CS, Greer AL, Fisman DN, Moghadas SM. Transmissibility of the 2009 H1N1 pandemic in remote and isolated Canadian communities: a modelling study. BMJ Open **2012**; 2: e001614.
9. Presanis AM, Pebody RG, Paterson BJ, et al. Changes in severity of 2009 pandemic A/H1N1 influenza in England: a Bayesian evidence synthesis. BMJ **2011**; 343:d5408.
10. Towers S, Feng Z. Pandemic H1N1 influenza: predicting the course of a pandemic and assessing the efficacy of the planned vaccination programme in the United States. Euro Surveill **2009**; 14:19358.
11. Reed C, Biggerstaff M, Finelli L, et al. Novel framework for assessing epidemiologic effects of influenza epidemics and pandemics. Emerg Infect Dis **2013**; 19:85–91.
12. Cintron-Arias A, Castillo-Chavez C, Bettencourt LM, Lloyd AL, Banks HT. The estimation of the effective reproductive number from disease outbreak data. Math Biosci Eng **2009**; 6:261–82.
13. Smith DJ, Lapedes AS, de Jong JC, et al. Mapping the antigenic and genetic evolution of influenza virus. Science **2004**; 305:371–6.
14. Pan K, Subieta KC, Deem MW. A novel sequence-based antigenic distance measure for H1N1, with application to vaccine effectiveness and the selection of vaccine strains. Protein Eng Des Sel **2011**; 24:291–9.
15. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature **2005**; 438:355–9.
16. Fraser C. Estimating individual and household reproduction numbers in an emerging epidemic. PLoS One **2007**; 2:e758.
17. Roberts MG, Nishiura H. Early estimation of the reproduction number in the presence of imported cases: pandemic influenza H1N1–2009 in New Zealand. PLoS One **2011**; 6:e17835.
18. Katriel G, Yaari R, Huppert A, Roll U, Stone L. Modelling the initial phase of an epidemic using incidence and infection network data: 2009 H1N1 pandemic in Israel as a case study. J R Soc Interface **2011**; 8:856–67.
19. Hedge J, Lycett SJ, Rambaut A. Real-time characterization of the molecular epidemiology of an influenza pandemic. Biol Lett **2013**; doi:10.1098/rsbl.2013.0331.
20. Edmunds WJ, Kafatos G, Wallinga J, Mossong JR. Mixing patterns and the spread of close-contact infectious diseases. Emerg Themes Epidemiol **2006**; 3:10.
21. Rohani P, Zhong X, King AA. Contact network structure explains the changing epidemiology of pertussis. Science **2010**; 330:982–5.
22. Nsoesie E, Mararthe M, Brownstein J. Forecasting peaks of seasonal influenza epidemics. PLoS Curr **2013**; doi:10.1371/currents.outbreaks.bb1e879a23137022ea79a8c508b030bc.
23. de Silva E, Ferguson NM, Fraser C. Inferring pandemic growth rates from sequence data. J R Soc Interface **2012**; 9:1797–808.
24. Mills CE, Robins JM, Lipsitch M. Transmissibility of 1918 pandemic influenza. Nature **2004**; 432:904–6.
25. Wu JT, Ma ES, Lee CK, et al. A serial cross-sectional serologic survey of 2009 Pandemic (H1N1) in Hong Kong: implications for future pandemic influenza surveillance. Influenza Other Respi Viruses **2011**; 5(Suppl 1): 190–4.
26. Wu JT, Cowling BJ. The use of mathematical models to inform influenza pandemic preparedness and response. Exp Biol Med (Maywood) **2011**; 236:955–61.
27. Ferguson NM, Cummings DA, Cauchemez S, et al. Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature **2005**; 437:209–14.
28. Cauchemez S, Boelle PY, Donnelly CA, et al. Real-time estimates in early detection of SARS. Emerg Infect Dis **2006**; 12:110–3.
29. Cauchemez S, Boelle PY, Thomas G, Valleron AJ. Estimating in real time the efficacy of measures to control emerging communicable diseases. Am J Epidemiol **2006**; 164:591–7.
30. Haydon DT, Chase-Topping M, Shaw DJ, et al. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. Proc Biol Sci **2003**; 270:121–7.
31. Wallinga J, Teunis P. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol **2004**; 160:509–16.
32. Wallinga J, Teunis P, Kretzschmar M. Reconstruction of measles dynamics in a vaccinated population. Vaccine **2003**; 21:2643–50.
33. Jombart T, Eggo RM, Dodd PJ, Balloux F. Reconstructing disease outbreaks from genetic data: a graph approach. Heredity (Edinb) **2011**; 106:383–90.
34. Cauchemez S, Valleron AJ, Boelle PY, Flahault A, Ferguson NM. Estimating the impact of school closure on influenza transmission from Sentinel data. Nature **2008**; 452:750–4.
35. Yang Y, Sugimoto JD, Halloran ME, et al. The transmissibility and control of pandemic influenza A (H1N1) virus. Science **2009**; 326:729–33.
36. Earn DJ, He D, Loeb MB, Fonseca K, Lee BE, Dushoff J. Effects of school closure on incidence of pandemic influenza in Alberta, Canada. Ann Intern Med **2012**; 156:173–81.
37. White LF, Pagano M. A likelihood-based method for real-time estimation of the serial interval and reproductive number of an epidemic. Stat Med **2008**; 27:2999–3016.
38. Wu JT, Leung GM, Lipsitch M, Cooper BS, Riley S. Hedging against antiviral resistance during the next influenza pandemic using small stockpiles of an alternative chemotherapy. PLoS Med **2009**; 6:e1000085.
39. Jin Z, Zhang J, Song LP, Sun GQ, Kan J, Zhu H. Modelling and analysis of influenza A (H1N1) on networks. BMC Public Health **2011**; 11 (Suppl 1): S9.
40. Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science **2004**; 303:327–32.
41. Koelle K, Cobey S, Grenfell B, Pascual M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. Science **2006**; 314:1898–903.
42. Donnelly CA, Finelli L, Cauchemez S, et al. Serial intervals and the temporal distribution of secondary infections within households of 2009 pandemic influenza A (H1N1): implications for influenza control recommendations. Clin Infect Dis **2011**; 52(Suppl 1): S123–30.
43. Xia Y, Gog JR, Grenfell BT. Semiparametric estimation of the duration of immunity from infectious disease time series: influenza as a case-study. J R Stat Soc Ser C Appl Stat **2005**; 54:659–72.
44. Balcan D, Goncalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. J Comput Sci **2010**; 1:132–45.
45. Van den Broeck W, Gioannini C, Goncalves B, Quaggiotto M, Colizza V, Vespignani A. The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. BMC Infect Dis **2011**; 11:37.
46. Bauch CT, Lloyd-Smith JO, Coffee MP, Galvani AP. Dynamically modeling SARS and other newly emerging respiratory illnesses: past, present, and future. Epidemiology **2005**; 16:791–801.
47. Johnson LF, White PJ. A review of mathematical models of HIV/AIDS interventions and their implications for policy. Sex Transm Infect **2011**; 87:629–34.