https://doi.org/10.1016/j.jtbi.2019.110070

Mechanistic Modelling of Multiple Waves in an Influenza Epidemic or Pandemic

Bo Xu [a, b], Jun Cai [a, b], Daihai He [c], Gerardo Chowell [d, e], Bing Xu [a, b, *]

[a] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China

[b] Joint Center for Global Change Studies, Beijing 100875, China

[c] Department of Applied Mathematics, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong (SAR), China

[d] Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, Georgia 30303, United States

[e] Division of International Epidemiology and Population Studies, Fogarty International Center, National Institute of Health, Bethesda, Maryland 20892, United States

[*] Corresponding author

Bo Xu, xu-b15@mails.tsinghua.edu.cn;

Jun Cai, cai-j12@mails.tsinghua.edu.cn;

Daihai He, daihai.he@polyu.edu.hk;

Gerardo Chowell, gchowell@gsu.edu;

Bing Xu, bingxu@tsinghua.edu.cn.

Abstract

Multiple-wave outbreaks have been documented for influenza pandemics particularly in the temperate zone, and occasionally for seasonal influenza epidemics in the tropical zone. The mechanisms shaping multiple-wave influenza outbreaks are diverse but are yet to be summarized in a systematic fashion. For this purpose, we described 12 distinct mechanistic models, among which five models were proposed for the first time, that support two waves of infection in a single influenza season, and classified them into five categories according to heterogeneities in host, pathogen, space, time and their combinations, respectively. To quantify the number of infection waves, we proposed three metrics that provide robust and intuitive results for real epidemics. Further, we performed sensitivity analyses on key parameters in each model and found that reducing the basic reproduction number or the transmission rate, limiting the addition of susceptible people who are to get the primary infection to infected areas, and limiting the probability of replenishment of people who are to be reinfected in the short term, could decrease the number of infection waves and clinical attack rate. Finally, we introduced a modelling framework to infer the mechanisms driving two-wave outbreaks. A better understanding of two-wave mechanisms could guide public health authorities to develop and implement preparedness plans and deploy control strategies.

Keywords: influenza outbreak; mechanistic model; multiple waves; number of infection waves; modelling framework

1. Introduction

In the temperate zone, consecutive waves of influenza infection have been documented for pandemics [1]. This multiple-wave pattern is one common feature that distinguishes influenza pandemics from epidemics [2]. For instance, the Spanish flu generated two waves in Geneva with the first in July and the second in October--November 1918 [3]; two waves in the US in the summer and autumn of 1918 respectively [4]. The Hong Kong flu caused two waves in Tristan da Cunha with the two epidemic peaks occurring in August--September in 1971 [5]. In 2009, three A/H1N1 waves have been documented in spring, summer and autumn in Mexico [6, 7], two waves peaking in July and October in Wales [8], and two waves peaking in June and November in Canada [9]. For annual seasonal influenza epidemics, two consecutive waves within the same influenza season have been recorded in some years in tropical areas, such as Hong Kong with waves occurring in March--May 2012 [10] and January--June 2015 [11].

Multiple waves of infection pose both challenges and opportunities [2]. On the one hand, successive waves may affect populations with unpredictable severity and transmissibility. On the other hand, the inter-wave period provides time for health authorities to prepare and respond, such as producing and delivering vaccines to high-risk individuals. For example, during the 2009 influenza pandemic in US, the number of cases during the second wave was larger than that during the first wave, with the peak of the second wave (week 42 in 2009) lagging behind that of the first

wave (week 24) by 18 weeks [12], and the vaccination program began in week 40 [13]. Besides, understanding the mechanisms driving the generation of multi-wave outbreaks could help health authorities develop and implement prevention and control strategies that prevent consecutive waves and mitigate severity.

An epidemic wave can be abstracted as a graph which plots the changing of incidence against time, and usually begins with a rapid rise to a peak and then falls more gradually [14]. To be specific, incidence refers to the number of new cases in a population generated during a time period [15]. Although the time scales of the multiple-wave outbreaks varied from several weeks [5] to months [16], some even exceeded years [17], the phenomenon discussed here is two waves within an influenza season (usually a single year), which is similar to the assumption in [18]. When we examine an epidemic curve that appears to exhibit multi-wave dynamics, it could be difficult to determine the precise number of waves without a quantitative criterion. To identify two-wave epidemics, Hoen *et al*. [19] defined a 2-peak (*TP*) metric. However, for epidemics of different sizes, the corresponding *TP* values vary, and the corresponding thresholds to distinguish between single-wave and multi-wave curves would be different. And the *TP* metric did not take into account the temporal information. Herein we propose three 2-wave metrics and five corresponding conditions, based on which the number of epidemic waves can be more precisely determined.

A number of research teams have assessed multiple-wave mechanisms, including reinfection by the same or another pathogen [5], non-pharmaceutical intervention measures [20] such as school closing and opening [21], spatial effects whereby the pathogen affects different segments of the population asynchronously [6], temporal variations in the transmission rate [22], the role of contact patterns [19], heterogeneous immunity patterns [2], and the synergistic interactions of multiple single causes [21, 23-25].

Among the existing mechanisms, human migration across locations as well as some man-made factors such as case reporting [3, 4] and vaccination behaviours [26] have been neglected as possible explanations for multiple-wave influenza outbreaks. Hence, a systematic analysis and a comprehensive classification system of potential explanatory models are necessary. The population level dynamics of human influenza epidemics stem from the interaction between the human population and influenza viruses during a period of time in specific locations. This complex process involves four dimensions: host, pathogen, space, and time. A small disturbance in a single dimension might break the homeostasis/equilibrium state of the epidemic and result in, for example, multiple waves within a year. Therefore, we summarize five new mechanistic models together with seven models previously described in literature. We then classify them into five categories according to host immune heterogeneity, virus strain heterogeneity, spatial scale and mobility, temporal variation of epidemiological parameters, and their combinations (Table 1).

When we use transmission models to simulate epidemics, the simulating results are linked to model structure, parameter values, initial conditions, and the intrinsic assumptions embedded within model formulations [27]. Thus, for a given model, it could be useful to map different model parameterizations to a distinct number of waves and infection severity of simulated epidemics. In [12], the number of initially infected and susceptible individuals, together with the time when they were reduced, had effects on the attack rate and the occurrence of the second wave. Boatto *et al*. [28] investigated the impacts of the average basic reproduction number (R_0), the number of initially immune individuals, the amplitude and period of the time-dependent transmission rate on the epidemic size over a year, and inferred the occurrence of epidemics, assuming the population size and the infectious period were fixed. Camacho and Cazelles [23] indicated the epidemic was more likely to be bimodal with a larger R_0. In [19], the epidemic size and the frequency of multi-wave epidemics were associated with R_0, but did not change with the infectious period. In above literatures, the severity of infection was usually measured by the attack rate or by the epidemic size. The attack rate refers to the proportion of the population that gets infected during a time period. It is more practical to use the "clinical" attack rate for infectious diseases like influenza. Clinical attack rate (CAR) is the proportion of the population who develops clinical symptoms after an infection [15], and can be calculated by dividing the number of newly infected symptomatic individuals by the number of people at risk of infection. One of the distinctions between the attack rate

and the epidemic size is that the former is a dimensionless parameter, with comparable values resulting from different initial population sizes. Hence in this paper we investigate the impact of different parameter values on the number of waves and the CAR in sensitivity analyses.

2. Methods

2.1. Definition of multi-wave epidemics

We define three two-wave metrics: Peak-Two metric (*PT*), Wave-Two metric (*WT*), and the time gap between two epidemic peaks ($Pgap$) (Eq. (1)-(3), Fig. 1), and five corresponding conditions (Table 2) to determine the number of epidemic waves. Three two-wave metrics are defined as follows.

$$WT(t_j) = \frac{y(t_j)-y(t_V)}{y(t_{P1})-y(t_V)} \quad (1)$$

$$PT = \frac{y(t_{P2})-y(t_V)}{y(t_{P2})} \quad (2)$$

$$Pgap = |t_{P1} - t_{P2}| \quad (3)$$

where $y(t)$ is the incidence time series. The subscripts 'P_1' and 'P_2' represent the highest peak and the second highest peak, respectively. t_{P1} and t_{P2} are the time points when the highest peak and the second highest peak arrive, respectively. t_j represents any time point except for t_{P1}. t_V satisfies the condition that $y(t_V) = \min y(t_k)$ with $t_{P1} < t_k \leq t_j \; or \; t_j \leq t_k < t_{P1}$, is the minimum value occurring in the time series between t_{P1} and t_j. The diagram of the three metrics are shown in Fig. 1.

Firstly determine $y(t_{P1})$, which is the maximum of $y(t)$. Secondly, in order to detect t_{P2}, calculate $WT(t_j)$ at all time points (t_j) except for t_{P1}. The time point corresponding to the maximum of $WT(t_j)$ is t_{P2}. Next, calculate three metrics: $WT(t_{P2})$, PT and $Pgap$. In particular, the definition of metric $WT(t_{P2})$ is similar to that of the "skip index" proposed in [29]. Then compare them with three thresholds: τ_1, $1/\tau_2$ and $\tau_3 * GT$, respectively, to determine the number of epidemic waves according to a set of criteria (Table 2). The number of waves can be one of the five values (1, 1.2, 1.5, 1.8, and 2), which is termed "fuzzy number" and represents the degree of truth that a specific epidemic is absolutely two-wave or one-wave. If the result of detection is 1.8 (or 1.2), the epidemic curve inclines to be bimodal (or unimodal). GT represents the average generation time of infection, which is defined as the period of time between the onset of the infectious period in a primary case to the onset of the infectious period in a secondary case infected by the primary case. In an epidemiological model without latent period (e.g. SIR model), GT can be approximated by the mean infectious period [30]. In an epidemiological model with latent period (e.g. SEIR model), GT refers to the sum of the average latent and infectious periods, when both periods follow the exponential distribution [7, 30-32], while another estimation of GT is the sum of the mean latent period and half the mean duration of infectiousness, with no specific restriction on their distributions [5, 33]. For simplicity, we choose the latter to calculate GT.

2.2. Mechanistic transmission models

2.2.1. Host-Immune-Heterogeneity

In the host dimension, temporal changes in an individual's immunity and variability in individual-level immunity can facilitate successive waves of infection. In this category, we include two transmission models: All-or-Nothing (AoN) and Partially-Protective-Immunity (PPI) from previous research [5, 16, 34]. After recovery from the prior infection, the AoN model assumes that a part of hosts are still susceptible to the same strain; the PPI model assumes that all hosts only develop incomplete protective immunity against the viral strain. The structural diagrams of the two models are shown in Fig. A.1(A) (Appendix).

2.2.2. Virus-Strain-Heterogeneity

In the pathogen dimension, the coexistence of two viral agents and virus mutation support bimodal epidemic curves. In this category, we adopt two models from previous research [5, 35, 36]: 2 Virus (2Vi) and Mutation (Mut), which assume that two different initiating strains exist and that a strain mutates into another strain within a host, respectively. The structural diagrams of the two models are shown in Fig. A.1(B-C) (Appendix).

2.2.3. Spatial-Scale-and-Mobility

In the spatial dimension, the spatial resolution of incidence data and the degree of connectivity across regions influence the likelihood of observing two-wave epidemics.

198 Thus, this category includes two models that integrate two unimodal epidemic curves
199 from two subregions, and consider the addition of susceptible people who come from
200 disease-free regions into infected regions, respectively (Fig. 2).
201
202 The 2 Region (2Reg) model we propose assumes a strain initiates an epidemic in a
203 subregion, and then triggers another epidemic in another subregion T_i days later.
204 When we analyze the incidence at the regional level, the two unimodal epidemic
205 curves in the two subregions are asynchronous. The aggregated incidence curve
206 across subregions could exhibit a bimodal shape (Fig. 2(A)).
207
208 The Importation-of-Susceptible (IoS) model we propose assumes a strain spreads
209 among people in a region, and a proportion (s_{at}) of susceptible people (S_t) at the time
210 t from other regions enter the region of interest during the downturn of the first wave
211 (Fig. 2(B)), which may subsequently increase the effective reproduction number and
212 promote a resurgence of the disease. The migration of people within a short period of
213 time such as mass gatherings can result from traditional festivals (e.g. the Spring
214 Festival in China, Thanksgiving Day in America), sporting events (e.g. the Olympic
215 Games, the World Cup), local armed conflicts and so on. For simplicity, we set two
216 adding proportions s_{a1} and s_{a2} at time T_S and $T_S + T_i$, respectively.
217
218 2.2.4. Temporal-Variation-of-Parameters
219 In the dimension of time, the temporal variation of epidemiological parameters can

partly explain the occurrence of two waves. Hence this category includes four models that incorporate time-varying reporting rate, transmission rate and susceptibility, respectively.

The Reporting-Rate-Variation (RRV) model we propose assumes the reporting rate fluctuates over a period of time (Fig. 2(C)). The reporting rate ρ_t is set to decrease firstly and then increase according to following piecewise linear function during a period of time, which is similar to the three-piece step function designed for the reporting rate in [29].

$$\rho_t = \rho_m - a\left(t - T_\rho\right),\ T_\rho < t \le T_\rho + \frac{1}{2}T_i$$

$$\rho_t = \rho_m - a\left(T_\rho + T_i - t\right),\ T_\rho + \frac{1}{2}T_i < t \le T_\rho + T_i$$

$$\rho_t = \rho_m,\ others \quad (4)$$

where T_ρ is the day when ρ_t begins to decline, T_i is the reporting change interval, t is time (day). ρ_m is the normal reporting rate outside T_i, as well as the maximum reporting rate in this model. a is the changing rate of ρ_t.

The Transmission-Rate-Variation (TRV) models assume the transmission rate is time-dependent. In this study, we adopt the "periodic transmission coefficient" model and improve the "derived time-dependent transmission coefficient" model proposed in [12], and refer to the first as Periodic-Transmission-Rate (PTR) and the second as Aperiodic-Transmission-Rate (ATR). The structural diagrams of the two models are shown in Fig. A.1(D) (Appendix). For the PTR model, the transmission rate β_t is set

to follow a periodic function with period of one year.

$$\beta_t = \beta_0 + \beta_1 \cos(2\pi t/365) \quad (5)$$

where β_0 is the average value of β_t, and β_1 is the amplitude of the fluctuating part.

For the ATR model, we improve the existing method used in [37] and [38], and calculated the non-negative aperiodic transmission rate β_t based on incidence $I(t)$, namely the number of new cases per unit time, rather than prevalence, on a time interval $[0, T_i]$, given infectious period $1/\nu > 0$ and latent period $1/\epsilon > 0$, in an SEIR model. The equations are as follows.

$$\beta_t = \frac{h(t)}{I(t)(S_0 - \int_0^{T_i} h(t)dt)} \quad (6)$$

$$h(t) = y'(t)/\epsilon + y(t) \quad (7)$$

$$I(t) = e^{-\nu t}(I_0 + \int_0^{T_i} e^{\nu t} y(t)\, dt) \quad (8)$$

where the incidence $y(t)$, initial fraction of susceptible (S_0) and infected (I_0) individuals, ϵ and ν should be given as the starting inputs of the algorithm. $y'(t)$ is the first-order derivative function of $y(t)$. Note that the following conditions should also be satisfied.

$$h(t) \geq 0, I(t) > 0, and\ S_0 > \int_0^{T_i} h(t)dt \quad (9)$$

The 2 Age-group (2Age) model we propose assumes a strain spreads across children and adults differently (Fig. 2(D)). The susceptibility of children decreases over time which may result from vaccination, whereas that of adults keeps stable. In this model, subscripts ‘C’ represents children. β_{ij} represents the effective contacts per day between the infectious people of class j and the susceptible people of class i. We

assume that no children mature and become adults within the study period of a single flu season.

$$s_{Ct} = a_1 t + b_1, \qquad 0 \le t < T_C$$

$$s_{Ct} = a_2 t + b_2, \qquad T_C \le t < T_C + T_i$$

$$s_{Ct} = a_2 (T_C + T_i) + b_2, \; t \ge T_C + T_i \tag{10}$$

where a_1 and a_2 are the changing rates of s_{Ct} during $[0, T_C)$ and $[T_C, T_C + T_i)$, respectively, b_1 and b_2 are the intercepts.

2.2.5. Combination

Each model discussed above represents a single-factor mechanism to explain the occurrence of two-wave epidemics. The combination of individual mechanisms may also be able to reproduce bimodal epidemic curves [39]. Two examples are provided as follows.

One was proposed by Chowell and colleagues to study the 1918 pandemic influenza in Geneva, Switzerland [3] and we call it 1918-Flu model. This model integrates RRV and TRV mechanisms and other factors concerning host heterogeneity of immune response, infectiousness/infectivity and infective period. The structural diagram of this model is shown in Fig. A.1(E) (Appendix).

The other is the Road-Network (RN) model proposed by us, which assumes a strain not only spreads among people within subpopulations but also travels to other

subpopulations with people carriers immigrating and emigrating via highway traffic (Fig. 2(E)). This model incorporates all the mechanisms described in the Spatial-Scale-and-Mobility category. In this model, the epidemic inside every single subpopulation can be modeled with an SIR model, while the interaction among subpopulations can be described by a transfer matrix $\mathbf{M}$ with the element m_{ij} representing the average number of individuals moving from subpopulation i to j per day. The transfer matrix is estimated using gravity model [40] with annual highway passenger transport volume and total population of each subpopulation, and spatial distance between subpopulations. Considering that the initial number of people who were actually susceptible to the virus and who would choose highway traffic for travel is not the total population (e.g. some elderly individuals owing pre-existing immunity [29]), and that in reality each individual interacts with a smaller group of individuals [41], we multiply the population in each subpopulation by a coefficient (θ) to calculate initial number of susceptible people. Another coefficient (δ) is multiplied by $\mathbf{M}$ to simulate the implementation of traffic control during the epidemic. All the subpopulations share the same parameters.

2.3. Numerical simulation

We assume that there is no birth or death, and the population are well mixed and homogeneous, ignoring population structure or social contact heterogeneity (except for the 2Age model), in all deterministic models. Model-simulated incidence is computed by counting the number of new hosts entering the infectious class per unit

time. Since it is difficult to report all attacks during surveillance and in order to take account of possible unreported asymptomatic cases, we assume the "observed" model-simulated incidence per unit time to be a random sample from a normal distribution with the mean $Inc \times \rho$, where Inc is the model-simulated incidence and ρ is the reporting rate. We simulate the models using the R package fitR [42].

2.4. Sensitivity analysis

We allow two or three epidemiological parameters in each model to vary within a certain range, with the values of other parameters being fixed as in Table A.1 (Appendix). For each set of parameter values, we use a certain model to produce a simulated incidence time series, and then calculate the CAR of this simulated epidemic and evaluate its number of infection waves using our proposed metrics and criteria. We repeat this process to investigate the impacts of different values of epidemiological parameters on the number of epidemic waves and the CAR.

3. Results

We use the parameter values in Table A.1 (Appendix) to simulate epidemic curves based on the 12 models, respectively (Fig. 3). In each model (except for the RRV, ATR, 1918-Flu, and RN model), we allow two or three epidemiological parameters (Table 3) to vary within a certain range, with the values of other parameters being fixed as in Table A.1 (Appendix), and investigate their corresponding impacts on the number of epidemic waves and the CAR. Note that all the numerical values for

parameters in this section are not correlated to any actual influenza epidemics or pandemics which has occurred in real life, but are from previous publications or directly given by ourselves. The presence of these parameter values is aimed to qualitatively explore how we can avoid the occurrence of successive infection waves and reduce the CAR by adjusting these epidemiological parameter values.

3.1. Host-Immune-Heterogeneity

Both the AoN model and the PPI model can simulate a bimodal epidemic curve with the first wave followed by a smaller second wave (Fig. 3(A-B)), using parameter values in Table A.1 (Appendix). Particularly, the second wave for the PPI model decays more slowly than that for the AoN model (Fig. 3(B)).

In the AoN model, for a given basic reproduction number R_0, the relationship between the CAR and the probability to develop long-term immunity α is approximately V-shaped (Fig. 4(A)). The points determined by the 2-tuples of (α,R_0) that correspond to the valley values of these V-shaped curves, can be approximated by an exponential curve (red, denoted as *C1*, R^2=0.948) (Fig. 4(D)). This model will not produce two-wave epidemics when R_0 <4, or when α >0.45 (Fig. 4(C)). The CAR is predominantly controlled by R_0. When R_0 takes larger values, the CAR will be higher (Fig. 4(A), (D)). In conclusion, if (α,R_0) follows *C1*, the epidemics follow one wave with locally lowest CAR.

In the PPI model, for a fixed R_0, the relationship between the CAR and the partial immune protection factor σ is approximately sigmoid, and a larger R_0 corresponds to a higher CAR (Fig. A.2(A)). Along a certain sigmoid curve in Fig. A.2(A), the point where the slope is the highest corresponds to a 2-tuple of (σ,R_0). The points determined by all these tuples can be approximated by a power function curve (red, denoted as *C2*, R^2=0.988) (Fig. A.2(D)). In Fig. A.2(C-D), we find when crossing *C2* from the lower right to the upper left, the CAR will decrease, and the resulting simulated curve is one-wave. Therefore, the epidemic can be unimodal with relatively low CAR, if (σ,R_0) is on the left side of curve *C2*. Fig. A.2 is in Appendix.

3.2. Virus--Strain-Heterogeneity

The 2Vi model can produce a bimodal curve using parameter values in Table A.1. Infections caused by strains 1 and 2 start at the same time, but each dominates the first and second wave, respectively (Fig. 3(C)). A two-wave epidemic occurs when basic reproduction numbers $R_0^1 \geq 1.6$ and $R_0^2 \geq 1.6$ (Fig. A.3(C)). Given $R_0^1 \leq 1.4$, the CAR increases little when R_0^2 increases substantially (Fig. A.3(A), (D)). For $R_0^1 > 1.4$, the relationship between the CAR and R_0^2 is approximately sigmoid, and a larger R_0^1 corresponds to a larger CAR (Fig. A.3(A)). Along a certain sigmoid curve in Fig. A.3(A), the point where the slope is the highest corresponds to a 2-tuple of (R_0^1,R_0^2). The points determined by all these tuples can be approximated by a power function curve (red, denoted as *C3*, R^2=0.850) (Fig. A.3(D)). In Fig. A.3(C-D), we find when crossing *C3* from the lower right to the upper left, the CAR will drop, and

the simulated epidemic will be unimodal. Hence, the epidemic can be one-wave with relatively low CAR, if (R_0^1,R_0^2) is on the left side of curve *C3*. Table A.1 and Fig. A.3 are in Appendix.

The Mut model produces a bimodal curve using parameter values in Table A.1. A strain initiates an epidemic wave that develops and decays gradually. At time T_m, the strain mutates into another strain, which infects additional people and causes another infection wave (Fig. 3(D)). Lower reproduction number R_0 and higher level of cross-immunity (1-σ) decrease the occurrence probability of two-wave epidemics (Fig. A.4(A)). In Fig. A.4(B), when σ is less than 0.10, as R_0 increases, the CAR rises when $R_0 < 4.2$, and drops when $R_0 > 4.2$. For a specific R_0, the CAR declines when σ increases. Table A.1 and Fig. A.4 are in Appendix.

3.3. Spatial-Scale-and-Mobility

The 2Reg model produces a two-wave epidemic using parameter values in Table A.1 (Appendix). A wave unfolds in subregion 1, and another wave in subregion 2 begins T_i days later (Fig. 3(E)). As expected, a lower R_0 can reduce the CAR. If T_i is short enough, the epidemic in each of the two subregions will overlap in time, resulting in a unimodal integrated curve.

The IoS model simulates a bimodal epidemic using parameter values in Table A.1. With the addition of susceptible people from other regions in day T_S and $T_S + T_i$, a

relatively small wave appears in the downward phase of the first wave (Fig. 3(F)). Given $T_S = 35$, lower values for both proportions (s_{a1},s_{a2}) of the added susceptible people will reduce the likelihood of two-wave epidemics (Fig. A.5(A)). Given $s_{a2} = 1.5$, the simulated epidemic will be one-wave when $T_S \notin [28, 57]$ (Fig. A.5(C)). Similarly, with $s_{a1} = 0.2$, the epidemic will be one-wave when $T_S \notin [27, 39]$ (Fig. A.5(E)). The CAR will drop when T_S is delayed (Fig. A.5(D), (F)), so it is preferable to delay T_S as much as possible. The second wave can be avoided by reducing s_{a1} or s_{a2} when T_S is within a certain range. Table A.1 and Fig. A.5 are in Appendix.

3.4. Temporal-Variation-of-Parameters

In the RRV model, the actual epidemic is unimodal, but the "observed" (being reported to the medical services) epidemic curve is bimodal, which results from oscillations of the reporting rate (Fig. 3(G)), with parameter values in Table A.1 (Appendix). We do not consider the situations when surveillance measures are taken to increase the reporting rate, because it will not influence the real CAR of the underlying epidemic.

For the PTR model, when transmission rate β_t reaches its minimum, the first wave begins to decline. As β_t increases, incidence goes up and reaches the peak of second wave when the slope of β_t curve reaches its maximum. Then β_t continues rising while incidence decreases (Fig. 3(H)), using parameter values in Table A.1. For a

specific β_1, the relationship between the CAR and β_0 is approximately V-shaped, except for abnormal β_0 that is smaller than β_1 (Fig. A.6(A)). Similarly, for a specific β_0, the relationship between the CAR and β_1 is V-shaped, except for abnormal β_1 that is larger than β_0 (Fig. A.6(B)). We identify three 2-tuples of (β_0,β_1) corresponding to the troughs of V-shaped curves, and related epidemic curves and transmission rates are displayed (Fig. A.6(C-D)). The points determined by the 2-tuples of (β_0,β_1) can be approximated by a linear regression line (green, denoted as *L1*, R^2=0.994, slope=1.29) (Fig. A.6(F)). The points determined by the tuples that correspond to bimodal curves are also distributed in a line (red, denoted as *L2*, R^2=0.834, slope=1.28), nearly parallel to line *L1* (Fig. A.6(E)). Note that β_0 must be no less than β_1, otherwise β_t would be negative. In conclusion, if (β_0,β_1) follows line *L1*, the epidemics can be unimodal with locally lowest CAR. Table A.1 and Fig. A.6 are in Appendix.

The ATR model incorporates all variability in transmissibility into β_t, which is derived from incidence data (here we use the weekly laboratory confirmed cases in US during the 2009 A/H1N1 pandemic, week 17-52 [12]). In Fig. 3(I), the simulated incidence changes asynchronously with β_t with a time lag of 2-3 days when comparing the peaks of both curves, using parameter values in Table A.1 (Appendix). The number of waves of the model-simulated epidemic curve and the corresponding CAR are determined by the incidence data and other epidemiological parameters as inputs to the model, as long as the conditions in Eq. (9) in the Methods section are

satisfied.

The 2Age model can simulate dual epidemic waves using parameter values in Table A.1. The first wave reaches its peak at the middle of the second stage of children's susceptibility (s_{Ct}). Shortly after s_{Ct} comes into the third stage (near zero), infections among adults dominate the second wave (Fig. 3(J)). β_{CC}, β_{AA}, and β_{CA} (or β_{AC}) represent the transmission rates among children, among adults, and between children and adults, respectively. Given $\beta_{CA} = 5/365$, the values of β_{CC} and β_{AA} which support two-wave epidemics are positively correlated (Pearson correlation coefficient is 0.756, *p-value*<0.001), and the values of β_{CC} are larger than those of β_{AA} (Fig. A.7(A)). Given $\beta_{AA} = 53/365$, (β_{CC}, β_{CA}) corresponding to two waves are shown in Fig. A.7(C), with $\beta_{CC} > 110/365$, and $\beta_{CA} < 10/365$. Given $\beta_{CC} = 170/365$, (β_{AA}, β_{CA}) that can produce dual waves are shown in Fig. A.7(E), with $\beta_{AA} \in [35/365, 70/365]$, and $\beta_{CA} < 9/365$. The CAR will decrease when β_{CC} is smaller, which is the same for β_{AA} (Fig. A.7(B), (D), (F)). In summary, the reduction of β_{CC} (or β_{AA}) with a moderate increase of β_{CA} could lower the CAR and the probability of two waves. Table A.1 and Fig. A.7 are in Appendix.

3.5. Combination

Using parameter values in Table A.1 (Appendix), the RN model yields a simulated two-wave epidemic in a given subpopulation (Fig. 3(K)). When adding together all the simulated epidemic curves in more than 300 subpopulations, we can get an

integrated curve whose estimated number of waves is 1.5 based on our proposed metrics and criteria (Fig. 3(L)). The multiple small peaks may result from the spatial coupling of subpopulations.

The 1918-Flu model simulated the transmission dynamics of the spring and autumn waves of the Spanish flu in Geneva, Switzerland, using parameter values in Table A.1 (Appendix). The first wave ranged from 1st July to 10th September, and the second started on 11th September (Fig. 3(M)).

4. Discussion

The annual number of deaths associated with influenza epidemics ranges approximately between ~300,000 and 650,000 globally [43]. This mortality burden results in heavy economic burden stemming from the damage to commerce and society [44], especially for pandemics which often display recurrent waves of infection during a short time period [2].

4.1. Robustness test of two-wave metrics and criteria

Firstly, we propose three metrics and a set of criteria to evaluate to what extent a given epidemic curve is bimodal or unimodal. To test their robustness, we apply them on real datasets including incidence time series of the 2009 pandemic in America [12] and several cities in mainland China, incidence time series of seasonal influenza A (H3N2) in Tristan da Cunha in 1971 [5], and of a Zika virus epidemic in a few Central

and South American countries in 2016 [45], respectively. It is shown that the fuzzy numbers of peaks detected by our metrics and criteria are consistent with visual inspection (Fig. A.8, Appendix). Therefore, in the sensitivity analysis, the numbers of infection waves of the simulated epidemic curves detected using these metrics are reliable.

4.2. Reality of summarized models

Secondly, to provide different explanations for the multi-wave dynamics, we explored 12 alternative mechanistic models classified into five categories, which are partly consistent with the strongest factors responsible for triggering waves of seasonal influenza [46]. Although some of these models seem to be manufactured, the corresponding scenarios have occurred in real life or have been reported in previous research.

4.2.1. Host-Immune-Heterogeneity

For an individual, her/his susceptibility to reinfection with a previously exposed strain would change with time, due to waning of immunity [25, 29] or lack of nutrition [39]. Among different individuals, the existence of the heterogeneity in their immunity against influenza viruses has also been reported by early works [5, 13, 47]. Therefore, the AoN and PPI models in this category are realistic.

However, it is noteworthy that reinfection is not common for interpandemic influenza

504 in the short term [48]. Reinfection to A/H3N2 was rare in each of the four annual
505 H3N2 epidemics from 2010 to 2014 in Hong Kong [10]. For pandemic influenza,
506 reinfection over the waves of the 1918 pandemic was also rare according to the
507 documents/data in thirteen English towns/schools [49]. The similar situation was for
508 the 2009 A/H1N1 pandemic [48]. A probable explanation is that the former wave
509 provided protection against infection during the latter wave [50]. Despite this,
510 studying reinfection can help in understanding the multi-wave epidemic patterns.
511

512 4.2.2. Virus-Strain-Heterogeneity

513 There are two examples of two-wave influenza epidemic in tropical regions, both of
514 which can be described by this model. One occurred in Hong Kong in 2012 [10], with
515 two waves mainly caused by influenza B and A/H3N2, respectively. The other
516 happened in Bangladesh in 2012 [51], with A/H1N1 dominating the first wave, and
517 subtype B the second. It is possible that the mainly affected age groups in two waves
518 were different, because different subtypes may prefer people of different ages. For
519 instance, the attack rate of A/H3N2 is higher for the elderly, and that of influenza B
520 towards youngsters is higher [11]. Besides, the emergence of mutated strains likely
521 caused different pathogenicity between waves, with various strains dominating
522 different periods of the 2009 pandemic [52]. Hence, the 2Vi and Mut models in this
523 category are reasonable.
524

525 4.2.3. Spatial-Scale-and-Mobility

2Reg It has been observed that the epidemic curve in a large geographical area is actually a composite of the curves from its constituent subareas [53] and can be termed synchronization/unsynchronization of neighboring subpopulations [54], e.g. the bimodal mortality incidence in some US cities during the Spanish flu [4, 20], the three waves in Mexico during the 2009 influenza pandemic [6]. Therefore, the 2Reg model has practical significance in real world and the spatial scale of the region under investigation should not be neglected.

IoS The increasing number of susceptible people, who are fuels for epidemics [20], could result from moving of people, who own no immunity against the pathogen, from disease-free regions into infected regions. For example, in a UK school in 1924, two batches of new healthy and susceptible boys were introduced into the school where influenza viruses were circulating, and subsequently multiple infection waves occurred [23]. Due to recent regional wars, displaced population with different levels of immunity, moving from war-affected regions to safe regions for survival, could accelerate the spatial spread of infectious diseases, e.g. cholera in Yemen with 3,000,000 displaced people [55], HIV in Ukraine with 1,700,000 internally displaced people [56]. Another interpretation of this model is the coming of new-born babies, who are fully susceptible to certain pathogens like measles virus. Besides, it may also cause a two-wave epidemic when exposed people who are infected but not infectious enter the region of interest. Thus the IoS model, which is similar to the "two-population model" in [18], is sensible.

4.2.4. Temporal-Variation-of-Parameters

RRV Changes in surveillance efforts and testing policies over time, which is related to case reporting, could affect the shape of epidemic curves [29, 55]. Besides spatial-variant reporting rates [57], time-variant reporting rates have also been observed during influenza pandemics [4, 29, 58], caused by many reasons, such as holidays or festivals [58], the definition of clinical cases and diagnostic criteria [57]. In this respect, the scenario described by the RRV model is realistic.

TRV Transmission rate is determined by population susceptibility, infectiousness and contact frequency of infectious individuals [3]. The time variation of transmission rate is related to the change of climatic conditions (e.g. temperature and absolute humidity [59]), human behaviours [26] (e.g. school cycle [6, 7], intervention measures [60]), and so on. The PTR model can simulate the above changing transmission rate. The original ATR model employed in [12] took the number of weekly reported cases as prevalence, which actually should be incidence, and estimated the aperiodic transmission rate β_t based on it. In practice, it is important to distinguish between prevalence and incidence [61]. Although [62] did not confound the two concepts, it deduced β_t from incidence in an SIR model, instead of the SEIR model used in our improved method.

2Age Due to the high susceptibility and incidence of school children during the

570 early phase of an epidemic [63, 64], children's initial susceptibility is set to be larger
571 than that of adults in our simulation, and children become a key target group of
572 vaccination [44], which is a reason for the decrease of their susceptibility [13] during
573 an epidemic. In real life, people determine whether to vaccinate themselves or their
574 children after considering the trade-off between side effects of a vaccine and its
575 benefits [26], so the vaccination coverage in a certain age group would fluctuate. This
576 is the reason why we set two different changing rates of children's susceptibility
577 during the first and second periods, respectively, in our model. Moreover, a survey
578 has revealed that a pattern that individuals tended to contact with people of the same
579 age was most pronounced in those aged 5–24 years, and less pronounced for older
580 people [64]. Therefore, in our simulation β_{CC} is set to be larger than β_{AA}, and much
581 larger than β_{CA}. The sensitivity analysis of this model reveals that the increase of β_{CC}
582 (or β_{AA}) together with a moderate reduction of β_{CA} could facilitate the occurrence of
583 two-wave epidemics. These are in line with the results in [19]. Therefore, the 2Age
584 model is practical.
585
586 4.2.5. Combination
587 The composite models incorporating multiple factors have been reported to reproduce
588 multi-wave influenza infections in UK [21, 25], US [13], and Mexico [24]. Another
589 research about Mexico [7] and a study on the epidemic arrival time of globally
590 spreading epidemics [18] developed meta-population models to explore travel effects,
591 which were similar to our proposed RN model. Hence, the 1918-Flu and RN models

592 in this category are realistic.

593

594 **4.3. Mitigation measures suggested by the sensitivity analysis**

595 Thirdly, in each of the above realistic models (except for the RRV, ATR, 1918-Flu,
596 and RN model), we conduct sensitivity analyses to investigate the effects of the
597 different values of the epidemiological parameters on the number of infection waves
598 and the CAR of simulated epidemics. For the target of fewer waves in a single
599 influenza season and a lower CAR, we summarize two categories of
600 recommendations on control measures from the results of the sensitivity analyses. The
601 first are measures aiming to reduce basic reproduction number or transmission rate.
602 The second are to limit the addition proportion of individuals who are to be primarily
603 infected, and to drop the probability of replenishment of people who are to be
604 reinfected.

605

606 **4.4. A modelling framework for investigating multi-wave epidemics**

607 In this research, we present a collection of possible mechanistic models to explain
608 two-wave influenza epidemic or pandemics occurring in a single season. Although
609 models can suggest plausible options, they cannot determine the actual mechanisms
610 [12]. Thus, fourthly, we summarize a modelling framework (Fig. 5) for choosing
611 appropriate models to reveal mechanisms underneath the bimodality of disease
612 surveillance data based on our proposed five categories of models.

613

At the very beginning, we would determine an appropriate scope of candidate single-factor mechanisms based on data itself, supplementary information and other sources of data regarding the epidemic and community to be analyzed. Firstly, we should pay attention to the spatio-temporal resolution of data. Different levels of administrative divisions correspond to different spatial scales. It is possible that during a specific period, an epidemic curve of two-wave pattern in a province is actually a summation of one-wave curves in cities belonging to the province, and vice versa [65]. In a specific location, a curve may be bimodal with a higher frequency of data sampling (e.g. one observation per week), and may be unimodal when the frequency is lower (e.g. one observation every ten days) (Fig. A.8(O-P), Appendix). Secondly, we could retrospect the information about the factors or events that would result in or accelerate the spatial migration of hosts, such as traditional festivals (e.g. Chinese Spring Festival, American Thanksgiving Day), sports events (e.g. Olympic Games), large conferences and exhibitions, long holidays (e.g. summer vacation, China's National Day), cross-region trade (e.g. live poultry trade). Thirdly, time-varying factors during the period of epidemic in certain specific locations should be considered, such as the ability or criteria of diagnosis and surveillance [66], personal hygiene habits (e.g. mask wearing, hands washing) and people's gathering behaviors in public places (e.g. schools, cinemas), measures or policies of intervention and control (e.g. border screening, vaccination, distribution of antiviral medicine) [52], extreme weather events (e.g. cold wave). Fourthly, other data sources would also provide insights about the epidemic. For example, by analyzing genetic

sequences [25] of sampled virus isolates, we are likely to figure out whether the original virus strain mutated to another strain or subtype, and whether there existed co-circulation of two or more strains during the epidemic; with serological data sampled from individuals who have ever been infected by the specific virus, we may test whether the titre of antibodies, which were produced by the immune system when the immune response was triggered by the virus infection, changed dramatically, and test whether a completely novel antibody appeared in vivo. Fifthly, if the above four steps do not supply enough information for us to select even one candidate single-factor mechanism, we can try to propose a hypothetical mechanism according to the specific research questions we concerned.

After the preliminary screening above, we might find only one factor associated with the two-wave epidemic in question and build a single-factor mechanistic model to explain it. But in real world, an epidemic is usually complex and influenced by more than one factors. Then we can assemble the related single-factor mechanisms in the scope and construct a composite mechanism model like those in Combination category. Next, we fit our model to the incidence data and derive particular parameter combinations that make the model outputs consistent with the epidemic. To evaluate the performance of the fitted model, in addition to the goodness-of-fit (e.g. AIC, R^2), a further important criterion is whether the parameter values inferred from a given model are biologically-reasonable [22], which should be judged according not only to previous researches and epidemiological investigation, but also to the information

from multiple data sources besides surveillance data.

Last but not least, suppose the variance of a certain parameter value is unusually large or the value itself is not biologically plausible, the corresponding factor would be excluded from the compounded mechanism model, and a new iteration of above steps can be conducted again.

4.5. Limitation and outlook

We have not been exhaustive as there may exist other possible mechanisms that are not covered in this work. The current categorization of mechanisms could be improved. For example, the IoS model has time-varying parameters, which could also be included in the Temporal-Variation-of-Parameters category. The transmission rate in the ATR model integrates all sources of variability during epidemic and it could be included in the Combination category. The 2Age model involves children's increasing immunity and could also be included in the Host-Immune-Heterogeneity category. Strictly speaking, the 2Reg model might not be a real mechanistic model, because the two-wave problem in a region is actually related to whether the epidemic data from two of its constituent subregions are combined, and the underlying epidemic in each subregion is still unimodal. In addition, the metrics and criteria proposed in this research to estimate the number of infection waves could be improved by taking into consideration the lasting time and size of each epidemic wave. All the simulations in this work are conducted on deterministic models, stochastic effects should be added in

680 the next step. In our proposed modelling framework, although fitting models to data
681 has been mentioned, we do not go through this in the current work. In further research,
682 the framework should be used to reveal the most reasonable mechanism to explain the
683 two infection waves of an actual influenza epidemic or pandemic with rigorous
684 statistical fits of models to real data.
685

686 ## Data accessibility

687 Code is available on GitHub via
688 https://github.com/BoXu123/multi-wave-influenza-outbreak.
689

690 ## Competing interests

691 We declare no competing interests.
692

693 ## Acknowledgements

694 Bo Xu, JC and Bing Xu were partially supported by the National Key Research and
695 Development Program of the Ministry of Science and Technology of the People's
696 Republic of China (2016YFA0600104), and donations from Delos Living LLC and
697 the Cyrus Tang Foundation to Tsinghua University. DH was partially supported by
698 the Early Career Schemes (PolyU 251001/14M) from Hong Kong Research Grants
699 Council. GC acknowledges support from NSF grant (1414374) as part of the joint
700 NSF-NIH-USDA Ecology and Evolution of Infectious Diseases program; UK
701 Biotechnology and Biological Sciences Research Council grant (BB/M008894/1) and

NSF grant (1318788 III). Bo Xu thanks Dr. Anton Camacho, Prof. Xiangjun Du, and Prof. Benjamin Cowling for helpful comments and suggestions.

References

[1] Fox, S.J., Miller, J.C. & Meyers, L.A. 2017 Seasonality in risk of pandemic influenza emergence. *PLoS Comp. Biol.* **13**, e1005749. (doi:10.1371/journal.pcbi.1005749).

[2] Miller, M.A., Viboud, C., Balinska, M. & Simonsen, L. 2009 The Signature Features of Influenza Pandemics — Implications for Policy. *New Engl. J. Med.* **360**, 2595-2598. (doi:10.1056/NEJMp0903906).

[3] Chowell, G., Ammon, C.E., Hengartner, N.W. & Hyman, J.M. 2006 Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. *J Theor Biol* **241**, 193-204. (doi:10.1016/j.jtbi.2005.11.026).

[4] Eggo, R.M., Cauchemez, S. & Ferguson, N.M. 2011 Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. *J R Soc Interface* **8**, 233-243. (doi:10.1098/rsif.2010.0216).

[5] Camacho, A., Ballesteros, S., Graham, A.L., Carrat, F., Ratmann, O. & Cazelles, B. 2011 Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proc. Biol. Sci.* **278**, 3635-3643. (doi:10.1098/rspb.2011.0300).

[6] Chowell, G., Echevarría-Zuno, S., Viboud, C., Simonsen, L., Tamerius, J., Miller, M.A. & Borja-Aburto, V.H. 2011 Characterizing the Epidemiology of the 2009 Influenza A/H1N1 Pandemic in Mexico. *PLoS Med.* **8**, e1000436. (doi:10.1371/journal.pmed.1000436).

[7] Tamerius, J., Viboud, C., Shaman, J. & Chowell, G. 2015 Impact of School Cycles and Environmental Forcing on the Timing of Pandemic Influenza Activity in Mexican States, May-December 2009. *PLoS Comp. Biol.* **11**, e1004337. (doi:10.1371/journal.pcbi.1004337).

[8] Keramarou, M., Cottrell, S., Evans, M.R., Moore, C., Stiff, R.E., Elliott, C., Thomas, D.R., Lyons, M. & Salmon, R.L. 2011 Two waves of pandemic influenza A(H1N1)2009 in Wales – the possible impact of media coverage on consultation rates, April – December 2009. *Eurosurveillance* **16**, 19772. (doi:10.2807/ese.16.03.19772-en).

[9] Earn, D.D., He, D., Loeb, M.B., Fonseca, K., Lee, B.E. & Dushoff, J. 2012 EFfects of school closure on incidence of pandemic influenza in alberta, canada. *Ann. Intern. Med.* **156**, 173-181. (doi:10.7326/0003-4819-156-3-201202070-00005).

[10] Wei, V.W.I., Wong, J.Y.T., Perera, R., Kwok, K.O., Fang, V.J., Barr, I.G., Peiris, J.S.M., Riley, S. & Cowling, B.J. 2018 Incidence of influenza A(H3N2) virus infections in Hong Kong in a longitudinal sero-epidemiological study, 2009-2015. *PLoS One* **13**, e0197504. (doi:10.1371/journal.pone.0197504).

[11] Tang, X., Fang, S., Chiu, A.P.Y., Lin, Q., Tang, E.Y.N., Wang, X. & He, D. 2018 Unsynchronized influenza epidemics in two neighboring subtropical cities. *Int J Infect Dis* **69**, 85-87. (doi:10.1016/j.ijid.2018.02.019).

[12] Mummert, A., Weiss, H., Long, L.P., Amigo, J.M. & Wan, X.F. 2013 A perspective on multiple waves of influenza pandemics. *PLoS One* **8**, e60343. (doi:10.1371/journal.pone.0060343).

[13] Towers, S. & Feng, Z. 2009 Pandemic H1N1 influenza: predicting the course of a pandemic and

742 assessing the efficacy of the planned vaccination programme in the United States. *Eurosurveillance* **14**,
743 19358. (doi:10.2807/ese.14.41.19358-en).
744 [14] The University of Liverpool, T.W.T. 2004 Understanding Epidemics Section 1: The Basics. In
745 *Understanding Epidemics* (pp. This website is developed as part of a project funded by the Wellcome
746 Trust which aims to improve the public understanding of epidemics, particularly their common
747 features, their spread (not every person during an epidemic will be affected), their impact (not
748 everyone infected will die), and their lasting influence (which will vary from disease to disease).
749 [15] Milwid, R., Steriu, A., Arino, J., Heffernan, J., Hyder, A., Schanzer, D., Gardner, E.,
750 Haworth-Brockman, M., Isfeld-Kiely, H., Langley, J.M., et al. 2016 Toward Standardizing a Lexicon of
751 Infectious Disease Modeling Terms. *Frontiers in Public Health* **4**. (doi:10.3389/fpubh.2016.00213).
752 [16] Mathews, J.D., McCaw, C.T., McVernon, J., McBryde, E.S. & McCaw, J.M. 2007 A biological model
753 for influenza transmission: pandemic planning implications of asymptomatic infection and immunity.
754 *PLoS One* **2**, e1220. (doi:10.1371/journal.pone.0001220).
755 [17] Wang, X., Jiang, H., Wu, P., Uyeki, T.M., Feng, L., Lai, S., Wang, L., Huo, X., Xu, K., Chen, E., et al.
756 2017 Epidemiology of avian influenza A H7N9 virus in human beings across five epidemics in mainland
757 China, 2013–17: an epidemiological study of laboratory-confirmed case series. *The Lancet Infectious*
758 *Diseases* **17**, 822-832. (doi:10.1016/s1473-3099(17)30323-7).
759 [18] Wang, L. & Wu, J.T. 2018 Characterizing the dynamics underlying global spread of epidemics.
760 *Nature Communications* **9**, 218. (doi:10.1038/s41467-017-02344-z).
761 [19] Hoen, A.G., Hladish, T.J., Eggo, R.M., Lenczner, M., Brownstein, J.S. & Meyers, L.A. 2015 Epidemic
762 Wave Dynamics Attributable to Urban Community Structure: A Theoretical Characterization of Disease
763 Transmission in a Large Network. *J. Med. Internet Res.* **17**, e169. (doi:10.2196/jmir.3720).
764 [20] Bootsma, M.C.J. & Ferguson, N.M. 2007 The effect of public health measures on the 1918
765 influenza pandemic in U.S. cities. *Proceedings of the National Academy of Sciences* **104**, 7588-7593.
766 (doi:10.1073/pnas.0611071104).
767 [21] He, D., Dushoff, J., Day, T., Ma, J. & Earn, D.J. 2013 Inferring the causes of the three waves of the
768 1918 influenza pandemic in England and Wales. *Proc. Biol. Sci.* **280**, 20131345.
769 (doi:10.1098/rspb.2013.1345).
770 [22] He, D., Dushoff, J., Day, T., Ma, J. & Earn, D.J.D. 2011 Mechanistic modelling of the three waves of
771 the 1918 influenza pandemic. *Theoretical Ecology* **4**, 283-288. (doi:10.1007/s12080-011-0123-3).
772 [23] Camacho, A. & Cazelles, B. 2013 Does homologous reinfection drive multiple-wave influenza
773 outbreaks? Accounting for immunodynamics in epidemiological models. *Epidemics* **5**, 187-196.
774 (doi:10.1016/j.epidem.2013.09.003).
775 [24] Herrera-Valdez, M.A., Cruz-Aponte, M. & Castillo-Chavez, C. 2011 Multiple outbreaks for the same
776 pandemic: Local transportation and social distancing explain the different "waves" of A-H1N1pdm
777 cases observed in Mexico during 2009. *Math Biosci Eng* **8**, 21-48. (doi:10.3934/mbe.2011.8.21).
778 [25] Dorigatti, I., Cauchemez, S. & Ferguson, N.M. 2013 Increased transmissibility explains the third
779 wave of infection by the 2009 H1N1 pandemic virus in England. *Proc Natl Acad Sci U S A* **110**,
780 13422-13427. (doi:10.1073/pnas.1303117110).
781 [26] Funk, S., Salathe, M. & Jansen, V.A. 2010 Modelling the influence of human behaviour on the
782 spread of infectious diseases: a review. *J R Soc Interface* **7**, 1247-1256. (doi:10.1098/rsif.2010.0142).
783 [27] Wearing, H.J., Rohani, P. & Keeling, M.J. 2005 Appropriate Models for the Management of
784 Infectious Diseases. *PLoS Med.* **2**, e174. (doi:10.1371/journal.pmed.0020174).
785 [28] Boatto, S., Bonnet, C., Cazelles, B. & Mazenc, F. 2017 SIR model with time dependent infectivity

parameter : approximating the epidemic attractor and the importance of the phase. In *Epidemics6*–*International Conference on Infectious Disease Dynamics* (Sitges, Spain.
[29] He, D., Lui, R., Wang, L., Tse, C.K., Yang, L. & Stone, L. 2015 Global Spatio-temporal Patterns of Influenza in the Post-pandemic Era. *Sci. Rep.* **5**, 11013. (doi:10.1038/srep11013 https://www.nature.com/articles/srep11013#supplementary-information).
[30] Wallinga, J. & Lipsitch, M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences* **274**, 599-604. (doi:10.1098/rspb.2006.3754).
[31] Anderson, R. & May, R. 1991 *Infectious Diseases of Humans. Dynamics and Control*. Oxford, Oxford Science Publications.
[32] Emilia, V. & Richard, G.W. 2010 *An introduction to infectious disease modelling*, Oxford University Press.
[33] Svensson, Å. 2007 A note on generation times in epidemic models. *Math. Biosci.* **208**, 300-311. (doi:https://doi.org/10.1016/j.mbs.2006.10.010).
[34] Gomes, M.G., White, L.J. & Medley, G.F. 2004 Infection, reinfection, and vaccination under suboptimal immune protection: epidemiological perspectives. *JTBio* **228**, 539-549.
[35] Mantle, J. & Tyrrell, D.A. 1973 An epidemic of influenza on Tristan da Cunha. *J. Hyg. (Lond.)* **71**, 89-95.
[36] Rios-Doria, D. & Chowell, G. 2009 Qualitative analysis of the level of cross-protection between epidemic waves of the 1918-1919 influenza pandemic. *J Theor Biol* **261**, 584-592. (doi:10.1016/j.jtbi.2009.08.020).
[37] Mummert, A. 2013 Studying the recovery procedure for the time-dependent transmission rate(s) in epidemic models. *J Math Biol* **67**, 483-507. (doi:10.1007/s00285-012-0558-1).
[38] Pollicott, M., Wang, H. & Weiss, H.H. 2012 Extracting the time-dependent transmission rate from infection data via solution of an inverse ODE problem. *J Biol Dyn* **6**, 509-523. (doi:10.1080/17513758.2011.645510).
[39] Lipsitch, M. & Viboud, C. 2009 Influenza seasonality: lifting the fog. *Proc Natl Acad Sci U S A* **106**, 3645-3646. (doi:10.1073/pnas.0900933106).
[40] Rodrigue, J.P., Comtois, C., Slack, B. 2013 *The Geography of Transport Systems*. London, Routledge.
[41] Keeling, M.J. & Grenfell, B.T. 2000 Individual-based perspectives on R(0). *J Theor Biol* **203**, 51-61. (doi:10.1006/jtbi.1999.1064).
[42] Camacho, A. & Funk, S. 2016 fitR: Tool box for fitting dynamic infectious disease models to time series. (R package version 0.1 ed.
[43] Iuliano, A.D., Roguski, K.M., Chang, H.H., Muscatello, D.J., Palekar, R., Tempia, S., Cohen, C., Gran, J.M., Schanzer, D. & Cowling, B.J. 2017 Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* **391**, 1285-1300. (doi:10.1016/S0140-6736(17)33293-2).
[44] Cox, R.J., Brokstad, K.A. & Ogra, P. 2004 Influenza virus: immunity and vaccination strategies. Comparison of the immune response to inactivated and live, attenuated influenza vaccines. *Scand J Immunol* **59**, 1-15. (doi:10.1111/j.0300-9475.2004.01382.x).
[45] Andersen, L. Digitized Zika cases and incidence rates by epidemiological week from PAHO. (
[46] Chattopadhyay, I., Kiciman, E., Elliott, J.W., Shaman, J.L. & Rzhetsky, A. 2018 Conjunction of factors triggering waves of seasonal influenza. *Elife* **7**, e30756. (doi:10.7554/eLife.30756).
[47] Hung, I.F.N., To, K.K.W., Lee, C.-K., Lin, C.-K., Chan, J.F.W., Tse, H., Cheng, V.C.C., Chen, H., Ho, P.-L.,

Tse, C.W.S., et al. 2010 Effect of Clinical and Virological Parameters on the Level of Neutralizing Antibody against Pandemic Influenza A Virus H1N1 2009. *Clin. Infect. Dis.* **51**, 274-279. (doi:10.1086/653940).

[48] Perez CM, F.M., Labarca JA. 2010 Pandemic (H1N1) 2009 reinfection, Chile. *Emerg Infect Dis* **16**, 156-157. (doi:10.3201/eid1601.091420).

[49] Health, M.o. 1920 Reports on public health and medical subjects No.4. In *Report on the Pandemic of Influenza, 1918-19* (London, UK, His Majesty's Stationery Office.

[50] Barry, J.M., Viboud, C. & Simonsen, L. 2008 Cross-protection between successive waves of the 1918-1919 influenza pandemic: epidemiological evidence from US Army camps and from Britain. *J Infect Dis* **198**, 1427-1434. (doi:10.1086/592454).

[51] Kelly, R. 2016 Hospital based influenza surveillance, Influenza seasonality graph, May 2007-January 2016. (FluTrackers.com.

[52] Yang, J.-R., Huang, Y.-P., Chang, F.-Y., Hsu, L.-C., Lin, Y.-C., Su, C.-H., Chen, P.-J., Wu, H.-S. & Liu, M.-T. 2011 New variants and age shift to high fatality groups contribute to severe successive waves in the 2009 influenza pandemic in Taiwan. *PloS one* **6**, e28288. (doi:10.1371/journal.pone.0028288).

[53] Cliff, A.D. & Haggett, P. 2006 A swash-backwash model of the single epidemic wave. *J Geogr Syst* **8**, 227-252. (doi:10.1007/s10109-006-0027-8).

[54] Balcan, D., Colizza, V., Goncalves, B., Hu, H., Ramasco, J.J. & Vespignani, A. 2009 Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc Natl Acad Sci U S A* **106**, 21484-21489. (doi:10.1073/pnas.0906910106).

[55] Camacho, A., Bouhenia, M., Alyusfi, R., Alkohlani, A., Naji, M.A.M., de Radiguès, X., Abubakar, A.M., Almoalmi, A., Seguin, C., Sagrado, M.J., et al. 2018 Cholera epidemic in Yemen, 2016–18: an analysis of surveillance data. *The Lancet Global Health* **6**, e680-e690. (doi:10.1016/s2214-109x(18)30230-4).

[56] Vasylyeva, T.I., Liulchuk, M., Friedman, S.R., Sazonova, I., Faria, N.R., Katzourakis, A., Babii, N., Scherbinska, A., Theze, J., Pybus, O.G., et al. 2018 Molecular epidemiology reveals the role of war in the spread of HIV in Ukraine. *Proc Natl Acad Sci U S A* **115**, 1051-1056. (doi:10.1073/pnas.1701447115).

[57] Riley, S. 2016 Epidemiology: Making high-res Zika maps. *Nat Microbiol* **1**, 16157. (doi:10.1038/nmicrobiol.2016.157).

[58] Yu, H., Cauchemez, S., Donnelly, C.A., Zhou, L., Feng, L., Xiang, N., Zheng, J., Ye, M., Huai, Y., Liao, Q., et al. 2012 Transmission dynamics, border entry screening, and school holidays during the 2009 influenza A (H1N1) pandemic, China. *Emerg Infect Dis* **18**, 758-766. (doi:10.3201/eid1805.110356).

[59] Deyle, E.R., Maher, M.C., Hernandez, R.D., Basu, S. & Sugihara, G. 2016 Global environmental drivers of influenza. *Proc Natl Acad Sci U S A* **113**, 13081-13086. (doi:10.1073/pnas.1607747113).

[60] McPake, B., Witter, S., Ssali, S., Wurie, H., Namakula, J. & Ssengooba, F. 2015 Ebola in the context of conflict affected states and health systems: case studies of Northern Uganda and Sierra Leone. *Confl Health* **9**, 23. (doi:10.1186/s13031-015-0052-7).

[61] Hadeler, K.P. 2011 Parameter estimation in epidemic models: simplified formulas. *Canadian Applied Mathematics Quarterly* **19**, 343-356.

[62] Hadeler, K.P. 2011 Parameter identification in epidemic models. *Math Biosci* **229**, 185-189. (doi:10.1016/j.mbs.2010.12.004).

[63] Ross, T., Zimmer, S., Burke, D., Crevar, C., Carter, D., Stark, J., Giles, B., Zimmerman, R., Ostroff, S. & Lee, B. 2010 Seroprevalence Following the Second Wave of Pandemic 2009 H1N1 Influenza. *PLoS Curr*

874 **2**, RRN1148. (doi:10.1371/currents.rrn1148).
875 [64] Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S.,
876 Tomba, G.S., Wallinga, J., et al. 2008 Social Contacts and Mixing Patterns Relevant to the Spread of
877 Infectious Diseases. *PLoS Med.* **5**, e74. (doi:10.1371/journal.pmed.0050074).
878 [65] Hirve, S., Newman, L.P., Paget, J., Azziz-Baumgartner, E., Fitzner, J., Bhat, N., Vandemaele, K. &
879 Zhang, W. 2016 Influenza Seasonality in the Tropics and Subtropics - When to Vaccinate? *PLoS One* **11**,
880 e0153003. (doi:10.1371/journal.pone.0153003).
881 [66] Lourenco, J. & Recker, M. 2014 The 2012 Madeira dengue outbreak: epidemiological
882 determinants and future epidemic potential. *PLoS Negl Trop Dis* **8**, e3083.
883 (doi:10.1371/journal.pntd.0003083).
884
885

Figure legends and Table captions

Fig. 1. Diagram of the definitions of three two-wave metrics. y is the number of new cases per day. The red points represent the highest peak (P_1), the second highest peak (P_2), and the local trough (V) between the time points of P_1 (t_{P1}) and j (t_j), respectively. t_j is any time point except for t_{P1}. The purple, green and yellow dashed lines are for metrics WT, PT, and $Pgap$, respectively. (A) when $t_{P1} < t_j$. (B) when $t_j < t_{P1}$, t_j coincides with t_V, and $y(t_j)$ coincides with $y(t_V)$.

Fig. 2. The structure diagrams of models proposed in this research. S, E, I, and R represent the number of people who are in four epidemiological states respectively: susceptible, exposed, infectious, and recovered/removed. Inc represents incidence, i.e. the number of new hosts entering the infectious class per unit time. β, ϵ, and ν are the transmission rate, the reciprocal of latent period, and the reciprocal of the infectious period, respectively. (A) 2 Region. Two dashed red arrows indicate that one infected individual arrives in subregion 1 and subregion 2 at time 0 and T_i, respectively. The subscripts '1' and '2' refer to two subregions, respectively. (B) Importation-of-Susceptible. S_t and s_{at} represent the number of susceptible people and the addition proportion of susceptible people at time t. (C) Reporting-Rate-Variation. ρ_t is the reporting rate. obs represents a proportion of incidence (Inc) and can be calculated by randomly sampling from a distribution (e.g. normal distribution) with the mean $Inc \times \rho_t$, because of the underreporting of cases. (D) 2 Age-group. The subscripts 'C' and 'A' refer to children and adults, respectively.

908 β_{ij} represents the effective contacts per day between the infectious people of class j
909 and the susceptible people of class i. s_{Ct} and s_A represent the susceptibility of
910 children at time t, and the susceptibility of adults, respectively. (E) Road-Network.
911 The subscripts '1', '2', '3', and '4' refer to four subpopulations. m_{ij} represents the
912 average number of individuals moving from subpopulation i to subpopulation j per
913 day.
914
915 **Fig. 3**. Number of daily new cases as a function of time in model (A) All-or-Nothing,
916 (B) Partially-Protective-Immunity, (C) 2 Virus (dotted lines represent incidence
917 caused by strain 1 and 2, respectively), (D) Mutation, (E) 2 Region (dashed lines
918 represent incidence occurred in subregion 1 and 2, respectively), (F)
919 Importation-of-Susceptible (s_{at} is the time-variant addition proportion of susceptible
920 people), (G) Reporting-Rate-Variation (ρ_t is the time-variant reporting rate), (H)
921 Periodic-Transmission-Rate (β_t is the time-variant transmission rate), (I)
922 Aperiodic-Transmission-Rate, (J) 2 Age-group (s_{Ct} is the time-variant susceptibility
923 of children), (K-L) Road-Network, and (M) 1918-Flu. The parameters in Table A.1
924 (Appendix) are used to produce the epidemic curves. The red line represents the
925 model-simulated daily incidence. The black line represents the "observed"
926 model-simulated daily incidence, which is a proportion of the model-simulated daily
927 incidence (Inc) and is calculated by randomly sampling from a normal distribution
928 with the mean $Inc \times \rho$, due to the underreporting of cases, with ρ being the case
929 reporting rate.

930

931 **Fig. 4**. Sensitivity analysis of the All-or-Nothing model. α is the probability to
932 develop long-term immunity, R_0 is the basic reproduction number. *PN* represents the
933 number epidemic waves. (A) Clinical attack rate as a function of α. Each black dot in
934 the curve indicates the point of the lowest attack rate. (B) Epidemic curves
935 corresponding to different R_0, given α=0.2. Numbers of epidemic waves (C) and the
936 attack rate (D) corresponding to different tuples of (α,R_0). Red line consists of points
937 of which each corresponds to a locally lowest attack rate along a certain V-shaped
938 curve of attack rate, which changes with α when R_0 is fixed.

939

940 **Fig. 5**. Flow chart of modelling framework.

941

942 **Table 1**. Characteristics of mechanistic transmission models that support multi-wave
943 epidemics.

944

945 **Table 2**. The criteria to determine the fuzzy number of epidemic waves (FNEW).

946

947 **Table 3**. The selected parameters and value ranges in each model for sensitivity
948 analysis.

949

Appendix. Supplemental Materials

Supplementary material associated with this article contains Table A.1 and Fig. A.1-Fig. A.8.

Fig. A.1. The structure diagrams of models adopted from previous research. S, E, I, R, and L represent the number of people who are in five epidemiological states respectively: susceptible, exposed, infectious, recovered/removed, and long-term protected. β, ϵ, ν, and γ are the transmission rate, the reciprocal of latent period, the reciprocal of the infectious period, and the reciprocal of temporary removed period, respectively. (A) All-or-Nothing (red) and Partially-Protective-Immunity (purple) [1]. (B, C) 2 Virus and Mutation [1]. Superscript stands for the infective strain, subscript for the already-immunized strain. Hosts recovered from strain i enter the L_i class and become completely protected against reinfection by strain i while remaining susceptible to the other circulating strain j. For the Mutation model, the two strains interact through a cross-immunity parameter $\sigma \in [0, 1]$ that acts by reducing the susceptibility to the other strain. The dashed arrow indicates that at time T_{mut} if $I^1 > 0$, one infectious host with the initial strain becomes infectious with the mutated strain. (D) Transmission-Rate-Variation (Periodic-Transmission-Rate and Aperiodic-Transmission-Rate) [2]. β_t is the transmission rate at time t. (E) 1918-Flu [3]. The subscript 'i' takes two values 1 and 2, referring to the first and second infection wave, respectively. N, J, and A represent the number of all people, hospitalized people, and asymptomatic people, respectively. μ, q, k, ρ, α, and γ are the natural birth (or death) rate, the relative infectiousness of asymptomatic people,

the proportion of clinically infections, the reporting rate, the diagnostic rate, and the reciprocal of the infectious period, respectively.

Fig. A.2. Sensitivity analysis of the Partially-Protective-Immunity model. R_0 is the basic reproduction number. 1-σ represents the degree of partial immune protection acquired after former recovery. PN represents the number epidemic waves. (A) Attack rate as a function of σ. Each black dot in the curve indicates the point of the highest slope. (B) Epidemic curves corresponding to different R_0, given σ=0.35. Numbers of epidemic waves (C) and attack rates (D) corresponding to different tuples of (σ,R_0). Red line consists of points of which each corresponds to the locally highest slope along a certain sigmoid curve of attack rate changing with σ when R_0 is fixed.

Fig. A.3. Sensitivity analysis of the 2 Virus model. R_0^1 and R_0^2 represent the basic reproduction number of virus 1 and virus 2, respectively. PN represents the number epidemic waves. (A) Attack rate as a function of R_0^2. Each black dot in the curve indicates the point of the highest slope. (B) Epidemic curves of two waves and one wave. Numbers of epidemic waves (C) and attack rates (D) corresponding to different tuples of (R_0^2,R_0^1). Red line consists of points of which each corresponds to a locally highest slope along a certain sigmoid curve of attack rate changing with R_0^2 when R_0^1 is fixed.

Fig. A.4. Sensitivity analysis of the Mutation model. R_0 is the basic reproduction

number of both virus strains. 1-σ represents the degree of cross-immunity acquired after recovering from the former virus strain. PN represents the number epidemic waves. Numbers of epidemic waves (a) and attack rates (b) corresponding to different tuples of (σ,R_0). (c) Epidemic curves corresponding to different R_0, given σ=0.36.

Fig. A.5. Sensitivity analysis of the Importation-of-Susceptible model. s_{a1} is the proportion of added susceptible individuals at the first time (T_S), while s_{a2} is that at the second time. Numbers of epidemic waves (A) and attack rates (B) corresponding to different tuples of (s_{a2},s_{a1}), given T_S=35. Numbers of epidemic waves (C) and attack rates (D) corresponding to different tuples of (T_S,s_{a1}), given s_{a2}=1.5. Numbers of epidemic waves (E) and attack rates (F) corresponding to different tuples of (T_S,s_{a2}), given s_{a1}=0.2.

Fig. A.6. Sensitivity analysis of the Periodic-Transmission-Rate model. β_0 is the average value of transmission rate (β) and β_1 is the amplitude of the fluctuating part of β. (A-B) Attack rate as a function of β_0 (A) and β_1 (B). Black dots are the troughs of V-shaped curves. The epidemic curves (C) and periodic transmission rates (D) both correspond to the black dots in (A, B). Numbers of epidemic waves (E) and attack rates (F) corresponding to different tuples of (β_0,β_1). Red points are determined by (β_0,β_1) corresponding to the locally lowest attack rates, and they could be approximated by a green line (*L1*). (β_0,β_1) corresponding to two infection waves would be approximated by a red line (*L2*). The white grids represent the meaningless

situation where $\beta_0 < \beta_1$.

Fig. A.7. Sensitivity analysis of the 2 Age-group model. β_{CC},β_{AA} and β_{CA} (or β_{AC}) represents the number of effective contacts per day among children, among adults, and between children and adults, respectively. Numbers of epidemic waves (A) and attack rates (B) corresponding to different tuples of (β_{AA},β_{CC}) , given $\beta_{CA} = 5/365$. Numbers of epidemic waves (C) and attack rates (D) corresponding to different tuples of (β_{CA},β_{CC}), given $\beta_{AA} = 53/365$. Numbers of epidemic waves (E) and attack rates (F) corresponding to different tuples of (β_{AC},β_{AA}), given $\beta_{CC} = 170/365$.

Fig. A.8. (A-N) The fuzzy numbers of peaks (PN) detected by the metrics and criteria proposed in our research when applying them on real datasets. These datasets are time series of reported cases of A (H3N2) in Tristan da Cunha in 1971 (A), of the 2009 pandemic in America (B) and several cities in mainland China (C-H), and of the Zika virus epidemic in a few Central and South American countries in 2016 (I-N). (O-P) The PN of the epidemic curve of A (H3N2) in Tristan da Cunha in 1971 corresponding to different temporal frequencies of data sampling: one value per week (O) and one value per ten days (P). The two blue dots in each epidemic curve are key points to determine PN of the specific curve using our method.

Table A.1. The parameters and corresponding values used in each model.

References in Appendix

[1] Camacho, A., Ballesteros, S., Graham, A.L., Carrat, F., Ratmann, O. & Cazelles, B. 2011 Explaining rapid reinfections in multiple-wave influenza outbreaks: Tristan da Cunha 1971 epidemic as a case study. *Proc. Biol. Sci.* **278**, 3635-3643. (doi:10.1098/rspb.2011.0300).

[2] Mummert, A., Weiss, H., Long, L.P., Amigo, J.M. & Wan, X.F. 2013 A perspective on multiple waves of influenza pandemics. *PLoS One* **8**, e60343. (doi:10.1371/journal.pone.0060343).

[3] Chowell, G., Ammon, C.E., Hengartner, N.W. & Hyman, J.M. 2006 Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. *J Theor Biol* **241**, 193-204. (doi:10.1016/j.jtbi.2005.11.026).