PLOS | COMPUTATIONAL BIOLOGY

The Spatial Resolution of Epidemic Peaks

Harriet L. Mills*, Steven Riley

MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

Abstract

The emergence of novel respiratory pathogens can challenge the capacity of key health care resources, such as intensive care units, that are constrained to serve only specific geographical populations. An ability to predict the magnitude and timing of peak incidence at the scale of a single large population would help to accurately assess the value of interventions designed to reduce that peak. However, current disease-dynamic theory does not provide a clear understanding of the relationship between: epidemic trajectories at the scale of interest (e.g. city); population mobility; and higher resolution spatial effects (e.g. transmission within small neighbourhoods). Here, we used a spatially-explicit stochastic meta-population model of arbitrary spatial resolution to determine the effect of resolution on model-derived epidemic trajectories. We simulated an influenza-like pathogen spreading across theoretical and actual population densities and varied our assumptions about mobility using Latin-Hypercube sampling. Even though, by design, cumulative attack rates were the same for all resolutions and mobilities, peak incidences were different. Clear thresholds existed for all tested populations, such that models with resolutions lower than the threshold substantially overestimated population-wide peak incidence. The effect of resolution was most important in populations which were of lower density and lower mobility. With the expectation of accurate spatial incidence datasets in the near future, our objective was to provide a framework for how to use these data correctly in a spatial meta-population model. Our results suggest that there is a fundamental spatial resolution for any pathogen-population pair. If underlying interactions between pathogens and spatially heterogeneous populations are represented at this resolution or higher, accurate predictions of peak incidence for city-scale epidemics are feasible.

Introduction

Novel respiratory pathogens continue to pose substantial public health challenges, not least because of the risk that large epidemics may overwhelm key health care resources such as vaccination stockpiles and intensive care facilities. Recent epidemics of concern include: SARS [1], influenza [2–4], H7N9 [5,6] and MERS [7,8]. During an epidemic it is important to accurately predict the impact of the epidemic over different spatial scales, where scale refers to the size of the region being monitored; such as a hospital, city, country or globally. Intervention policies should be defined relative to this spatial scale, for example taking account of how long it will take to vaccinate a whole city or to distribute a treatment country-wide. Those making decisions about intervention strategies need a clear understanding of the underlying epidemic process, so as to anticipate the magnitude and timing of peak incidence at their scale of interest and to effectively control the epidemic.

Spatially explicit transmission models are used frequently to increase understanding of the spread of epidemics caused by pathogens which transmit between individuals close in space. For example: influenza [9–11], measles [12–14], and smallpox [15,16] have all been represented by spatially explicit epidemic models. All of these examples can be thought of as metapopulation models in which the population of interest is represented as a collection of sub-populations located in space, for example households [17–19], airports (GLEaM [20]) or districts/states [21]. The advantages of these models are that they can capture complicated mobility and mixing patterns and heterogeneous population density, without the complexity of an individual-based model. Also, model output can be easily reported for specific populations, such as counties or cities.

It is known that heterogeneity both in population density and typical mixing behaviour heavily influence disease spread. Both of these are defined according to the resolution of the population representation, where resolution defines the number and size of the pixels making up the "image" of the population within the model. A pixel is the smallest single component of an image. A high resolution representation will divide the region into many small pixels; a lower resolution uses fewer, larger pixels (Fig. S1). The resolution chosen is usually decided by the data available: for example, population and travel data may be defined at the ward or county level only. The level of mixing between individuals in distinct pixels is defined by mobility models, these are often fitted to travel data from censuses. Also, sometimes, resolution is limited by computational capacity.

The concepts in our paper require precise definitions of the terms: scale, resolution and pixel. The literature using these three words is somewhat ambiguous with the terms resolution and scale sometimes used interchangeably. Therefore, for clarity, we have

Author Summary

Fundamental spatial processes such as individuals' interactions and movement are not sufficiently well understood and yet they define the transmission of infectious diseases through populations. Spatial models of epidemics represent the region of interest (such as a city or country) as a collection of spatial units. To anticipate the magnitude and timing of peak incidence and to predict demand on health care resources in the region a clear understanding is needed of the relationship between the resolution of the representation (number and size of the pixels), the population interactions and the epidemic trajectories. We used a spatially explicit meta-population model of disease transmission to demonstrate that thresholds existed such that models with too low a resolution overestimated peak incidence, implying that ill-defined models may result in incorrect predictions. However, the results suggest that if population interactions are represented in sufficient detail, accurate estimates of peak demands on key health care resources are feasible.

included explicit definitions at first use of the words (above) and in Table 1.

Results

We implemented a generic metapopulation model with arbitrary spatial resolution (see Methods) varying from approximately 1km^2 ($30''$ by $30''$, the smallest unit representation) upwards (Fig. S1). We generated a theoretical population density in a region with total population just over 4 million and of size approximately 49km^2 (49×49 pixels). The region had three 'urban' areas where population density was generated using a 2-dimensional bivariate Gaussian and a 'rural' area, generated from a uniform distribution, Fig. 1G. We used this formulation to simulate the spread of a pathogen representative of influenza, with an SIR-like natural history, assuming that the generation time was 2.6 days and the basic reproductive number R_0 was 1.8. The epidemic was seeded with 10 individuals in a central region (Fig. 1G), simulations were repeated 25 times at each resolution.

The within-pixel contact rate was fixed for all pixels. Mobility between pixels was represented by a kernel with an offset power function. A kernel defines the relative probability of travelling between two pixels. The offset power function is an adaptation of the gravity model. The gravity model states that an individual's probability of mixing in a pixel different to their home pixel is inversely proportional to the distance apart of the pixels, to some power. The offset power function adds in an offset distance parameter, which means that pixels closer together than this distance mix fully. See Methods for a full definition of the kernel and the resulting mobility model.

Initially we considered three different kernels: we used an offset of 2 km and three different powers giving low, medium and high contact between pixels (the power, γ, was -6, -4 and -2 respectively), Fig. S2. The highest mobility kernel is in line with kernels fitted to commuter data in the UK and US [9]. However, our review of data on travel patterns found that only 15% of an average individual's journeys are commuting, making up just 19% of the total distance an average individual travels each year [22] (Table S1). Commuting data also excludes key at-risk groups – the under 17 s and over 70 s – who have lower mobility travel patterns compared to the 18–69 population [22]. Therefore, we explored more restrictive kernels than those estimated using commuting data to reflect shorter distances travelled and lower frequency travel in the most at-risk populations and the regular non-commuting travel of the wider population.

We confirmed that the overall cumulative attack rate (CAR) for our model was independent of the mobility kernel and the model resolution (Fig. 1A–F). This was by design: the model was constructed such that with the assumption of mass action mobility (the rate of contact between two groups is proportional to the size of each of the groups) the epidemic was identical at every resolution. This means that the next generation matrix at any resolution and for any mobility has the same spectral radius: R_0 was the same at all resolutions and contact levels and the local and global R_0s were the same. A full proof that R_0 was constant with respect to resolution is given in the Text S1 in Supporting Information S1, and is similar to that in Ref [23]. Because R_0 was constant, if the mobility was such that there was contact between every pair of pixels, the final epidemic size was the same across all resolutions and in every pixel. If mobility was restrictive enough that some pixels were never infected the final CAR reflected this restriction. The full proof that attack rates were constant with respect to resolution is in the Text S2 in Supporting Information S1, and is similar to those in Refs [24,25].

For the theoretical population density, the existence of a fundamental spatial resolution was apparent: at resolutions lower than this threshold, system-wide peak incidence was substantially over-estimated, obtaining a high peak incidence and fast spread similar to that obtained in a fully mixed model (the lowest resolution). However, at the fundamental resolution and above, consistent estimates of the peak attack rate were obtained (Fig. 1H). This was increasingly evident as mobility became more and more restricted: for the most localised mobility assumptions (low power), peak incidence in the fully mixed case was nearly double that at the highest resolution. At high resolutions, multiple small pixels containing low numbers of individuals and with a high heterogeneity in population size slowed the epidemic spread; resulting in a long epidemic duration and a low peak incidence compared to low resolution model scenarios.

Increased mobility reduced the effect of resolution on the epidemic trajectory. At medium mobility, peak incidence increased with decreasing resolution but there was no distinct

Table 1. Explicit definitions of the terminology used in this work.

Word	Definition
Scale	The relative size or extent of a region.
Resolution	The degree of detail visible in an image - in our work this defines the number and size of the pixels making up the representation of the region.
Pixel	The smallest single component of an image (or in our model, the smallest single component of the population representation).

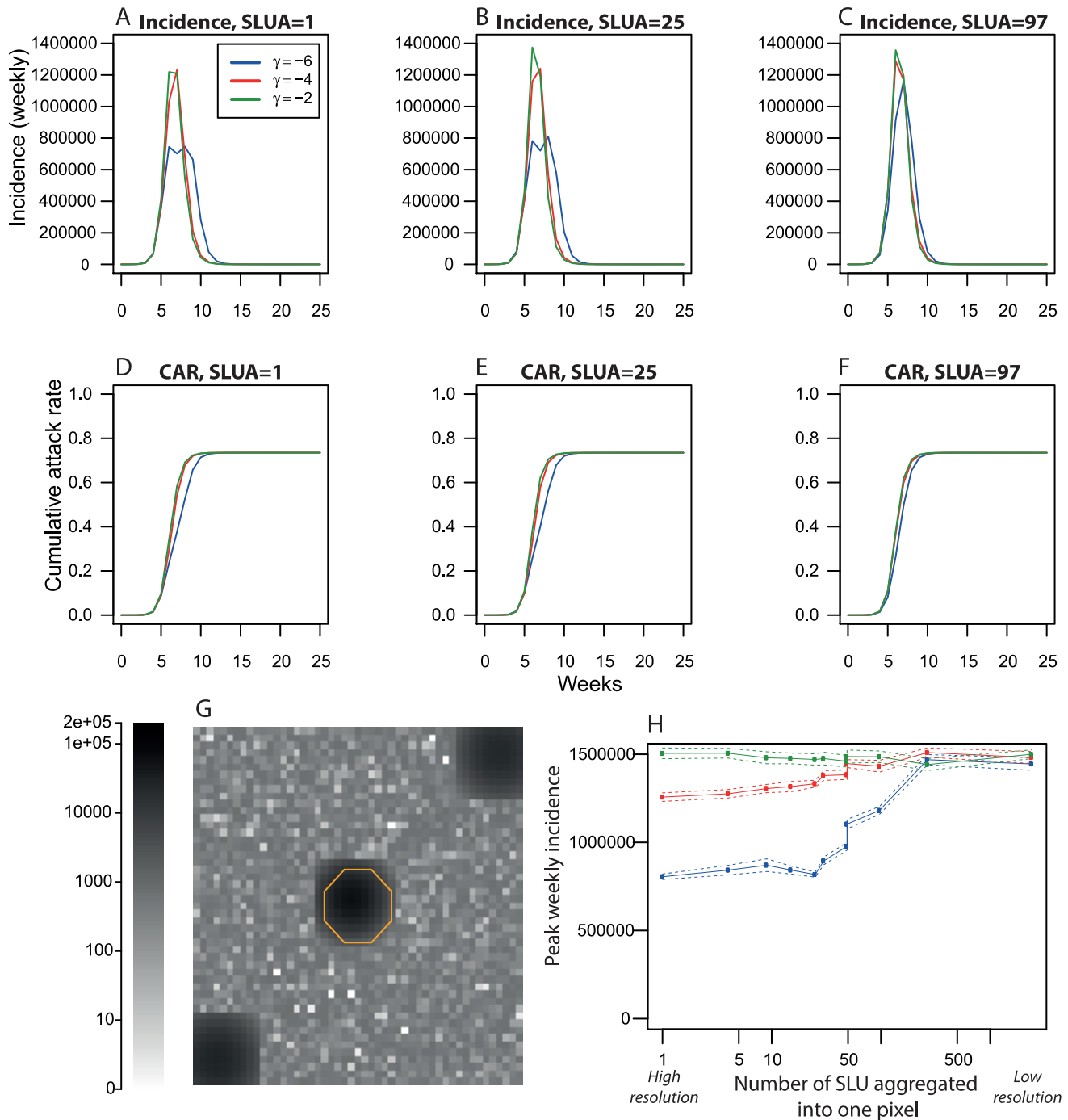doi:10.1371/journal.pcbi.1003561.t001

Figure 1. The effect of resolution and mobility in a theoretical population. A–C: Weekly incidence with time, for different spatial resolutions of the population. SLUA is the number of smallest LandScan units aggregated to make one pixel at that resolution. SLUA = 1 is the highest resolution possible in the data. D–E: Cumulative Attack Rate (CAR) for different resolutions of the population. Note that CAR is the same for all mobility levels and for all resolutions, but the epidemic trajectories differ. G: The population density of the theoretical population, the epidemic was seeded with 10 cases placed at random in the area indicated by the orange outline. H: Peak incidence for different levels of mobility (blue, very restrictive, green least restrictive). The number of smallest LandScan units aggregated to make one pixel increases from left to right along the x-axis (and resolution decreases). The spatial spread of the epidemic in these three scenarios is illustrated in a Movie S1.
doi:10.1371/journal.pcbi.1003561.g001

threshold. At the highest mobility, peak incidence was unaffected by resolution: the high level of contact between pixels facilitated the quick spread of the epidemic, indicated by a short epidemic duration and a high peak incidence at every resolution (Fig. 1A–C).

Resolution and mobility remained important when the model was constructed with real population densities. We repeated the analysis (using the same three kernels) for four regions selected from LandScan data [26]: Guangzhou, Rio de Janeiro, Delhi and New York (Fig. 2A). The smallest LandScan unit is approximately

1 km^2 (30″ by 30″) in size. The effect of resolution was most evident when mobility was more restricted, as with the theoretical population. In Guangzhou, Rio and New York, changing the spatial resolution had a significant effect on the peak incidence when mobility was at a low to medium level, though the effect was less clear in Delhi (Fig. 2B–E). The Delhi region had the largest total population size and the highest mean population density of all regions we considered (Table S2 and Fig. S3). Therefore, even at low mobility the numbers mixing will be relatively high, meaning that the disease spread will not be as restricted as it would be in a less densely populated region.

We used Latin Hypercube Sampling (LHS) [27] to determine whether the patterns we saw with the illustrative mobility kernels could be generalised within a wider parameter space of mobility functions. We varied the kernel parameters: the power, γ, between -6 and -2 (as discussed earlier, this selection gave a wide range of mobility levels) and the saturation distance between 1 and 10 km, choosing from a log scale (so smaller distances are more likely). We tested 50 parameter sets chosen using the LHS technique [27], with 10 separate realisations of each set for each region (variation in results from the stochastic model was low - see confidence intervals for 25 repeats in Fig. 2 for example). Kernels for the 50 sets are plotted in Fig. S4. The LHS results confirmed that the effect of resolution is most important in populations which are less mobile, Fig. 3. As mobility decreased (a combination of the offset and the power in the kernel) the difference in peak incidence between the lowest and the highest resolution increased. This was particularly true in Guangzhou, Rio and New York, but in Delhi the effect was reduced (due to Delhi having a very large population in comparison to the other regions).

Recently it has been suggested that the movement of individuals depends not only on the source and destination cities, but also on the population density of the surrounding area [28]. This model is called the radiation model and has been proposed as a distinct alternative to the gravity model. However, we calculated the actual number of individuals moving between pixels (the flux) and found the radiation model flux to be very close to the offset gravity model of medium mobility, particularly at the highest resolution, Fig. 4. Indeed the radiation model is always bounded by the three gravity models we use and our LHS models explore a large space around these. More generally, gravity-like models have been implemented with a number of different normalisation assumptions, some of which produce population flux patterns very similar to the radiation model [15,17].

Discussion

When managing epidemics it is desirable to know the size and duration of the epidemic and the magnitude and timing of the peak incidence over the spatial scale of interest [4,29,30]. This scale of interest may be a city, a region or a whole country. Resources such as treatment, vaccinations and diagnostic tests will take time to be deployed over this scale and it can take time to develop and generate enough of these resources for the whole affected population [31,32]. Accurate predictions about the magnitude and timing of peak incidence would greatly enhance the ability of public health officials to effectively limit the impact of epidemics.

We have shown how the representation of population interactions can impact model estimates of key epidemic outcomes. We examined the effect of the resolution of the population density on the model predictions of epidemic spread over the scale of interest. We refer to resolution as defining the number and size of the individual pixels dividing the region; higher resolution representations use a higher number of smaller pixels. Our results imply that for plausible population densities and mobility patterns, fundamental resolutions exist for specific pathogens such that the detail of the population and their interactions must be represented faithfully if accurate epidemic trajectories are to be estimated.

The impact of model resolution was clear in models of less mobile populations: our results indicate that at lower mobility, low resolution representations overestimated the peak incidence, obtaining a high peak incidence and fast spread similar to that obtained in a fully mixed model. However, sufficiently high resolution representations gave lower and later peak incidences because of the delaying effect of multiple small pixels. Indeed at low mobility, clear thresholds existed for the resolution of the theoretical population density, such that models with resolutions below the threshold over-estimated the system-wide peak incidence. Similar thresholds existed for real population densities: Guangzhou, Rio, New York and Delhi. Increased mobility reduced the effect of resolution on the epidemic.

The kernels which were most affected by resolution were those which gave a lower mobility than that identified by commuting data (Table S1). Generally children are considered to cause the majority of transmission of pathogens like flu and measles, because their level of age group assortative mixing is very high [33,34]. Children also travel less far than working adults [22]. Together, these imply that a kernel for children is likely to be more restrictive than those defined by commuting data alone. Therefore, our results indicate that the correct specification of population interactions and sufficient spatial resolution is particularly relevant for epidemics such as measles and flu - those in which children play a large role.

Although we have considered age effects implicitly by including lower mobility levels than are reported for commuting data (Table S1), the explicit representation of age within a similar modelling framework may lead to additional insight. For example, transmission dynamics at different scales may be driven by different age groups: the behaviour of more mobile adults may be disproportionately important in the seeding of nearby pixels. However, the slower than expected within-country spatial spread during 2009 [35] suggests that for pandemic influenza, population sub-groups with reduced mobility likely do define the fundamental resolution.

We have chosen to represent the real biological process by a high resolution metapopulation model. Although we have not been able to push the model to resolutions higher than 1 km by 1 km, we suggest it is reasonable to assume, for the mobility kernels considered here, that the thresholds observed for peak incidence would not change substantially were we to approach the resolution of an individual-based model.

The model used here was intended specifically to test only the changing resolution of the disease transmission process. By design we did not want to assume that transmissibility was intrinsically higher or lower in different parts of the population. In future work, we hope to calibrate this model structure using actual disease incidence data and (after a minor modification to the definition of the force of infection) test for the possibility that population density affects transmissibility.

Although it is somewhat reassuring that estimates of peak incidence are biased upwards if resolution is too low, the epidemic duration is underestimated. In order to avoid the effects of incorrect model specification, where possible, spatial resolution should be treated in a similar manner to temporal resolution in fixed-time-step models: neither the doubling nor halving of spatial resolution should have a substantive effect on key model outputs.
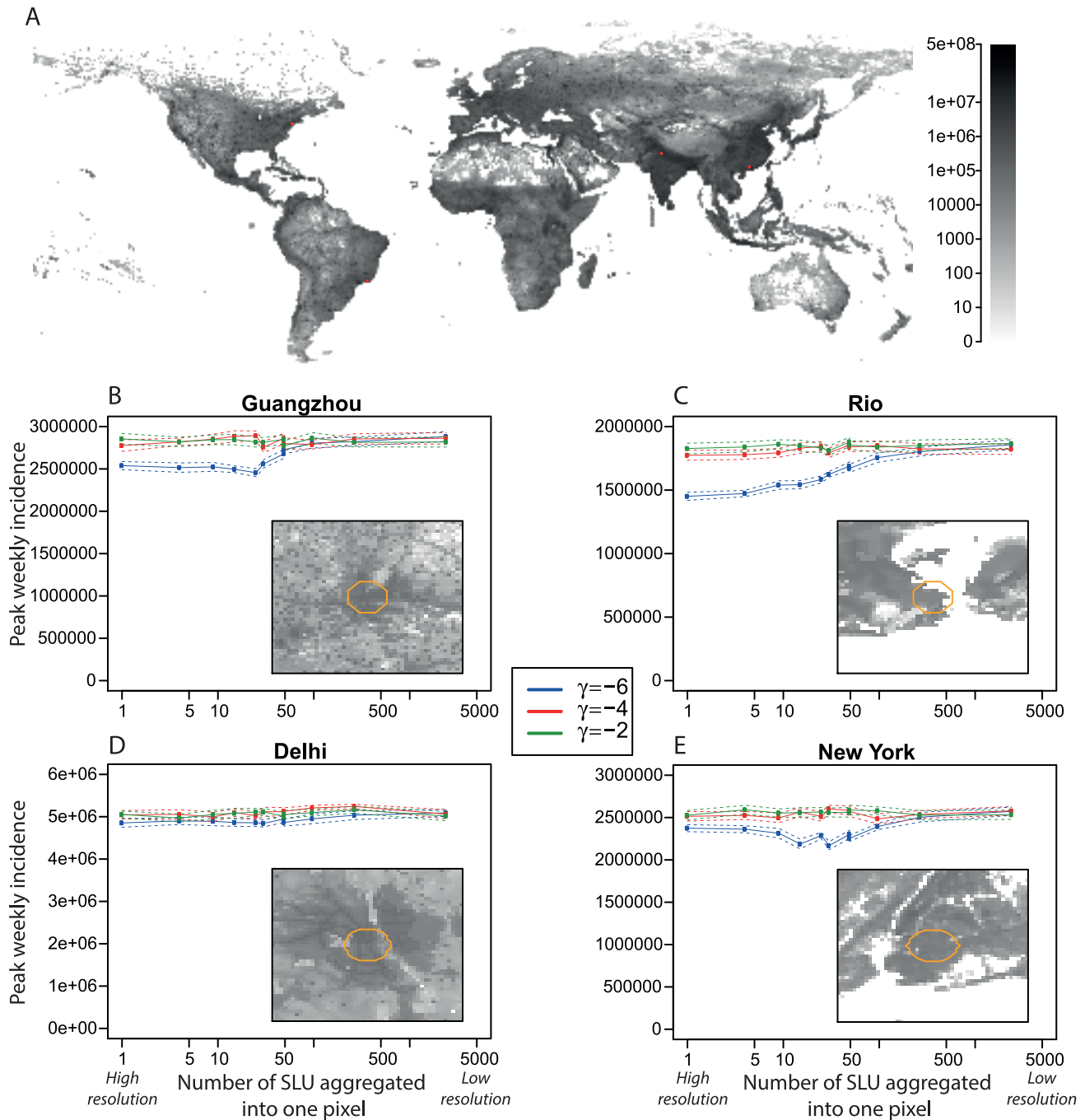
Figure 2. Results for actual populations from LandScan data [**26**]. SLU are the Smallest Landscan Units (30″ by 30″, or approximately $1km^2$). A: Population density of the whole world, the red spots indicate the locations of the 4 population densities we considered. The regions were chosen as a 49×49 unit square around their official lat-long centre. B–D: Main graphs indicate the peak incidence at each resolution, for Guangzhou (B), Rio (C), Delhi (D) and New York (E). The insets are the population densities of these regions, the colour scale is the same as panel A. The epidemics were seeded in a circle of radius 5 km around the official lat-long centre of the regions, the possible seeding area is marked by an orange outline on the insets.

doi:10.1371/journal.pcbi.1003561.g002

Methods

Model description

We defined a spatially explicit meta-population model as follows (similar to Ref. [21]). A given region of known population density was represented as D pixels, such that each pixel (index i) is the same spatial size but the number of individuals in the pixel (N_i) varied according to location. Mixing between and within each pixel was determined by a mobility model, represented by a matrix m such that an entry m_{ij} was equal to the probability that for an individual from pixel i, given that the individual made a contact, this contact was with an individual from pixel j (mobility was defined using a kernel, discussed later).
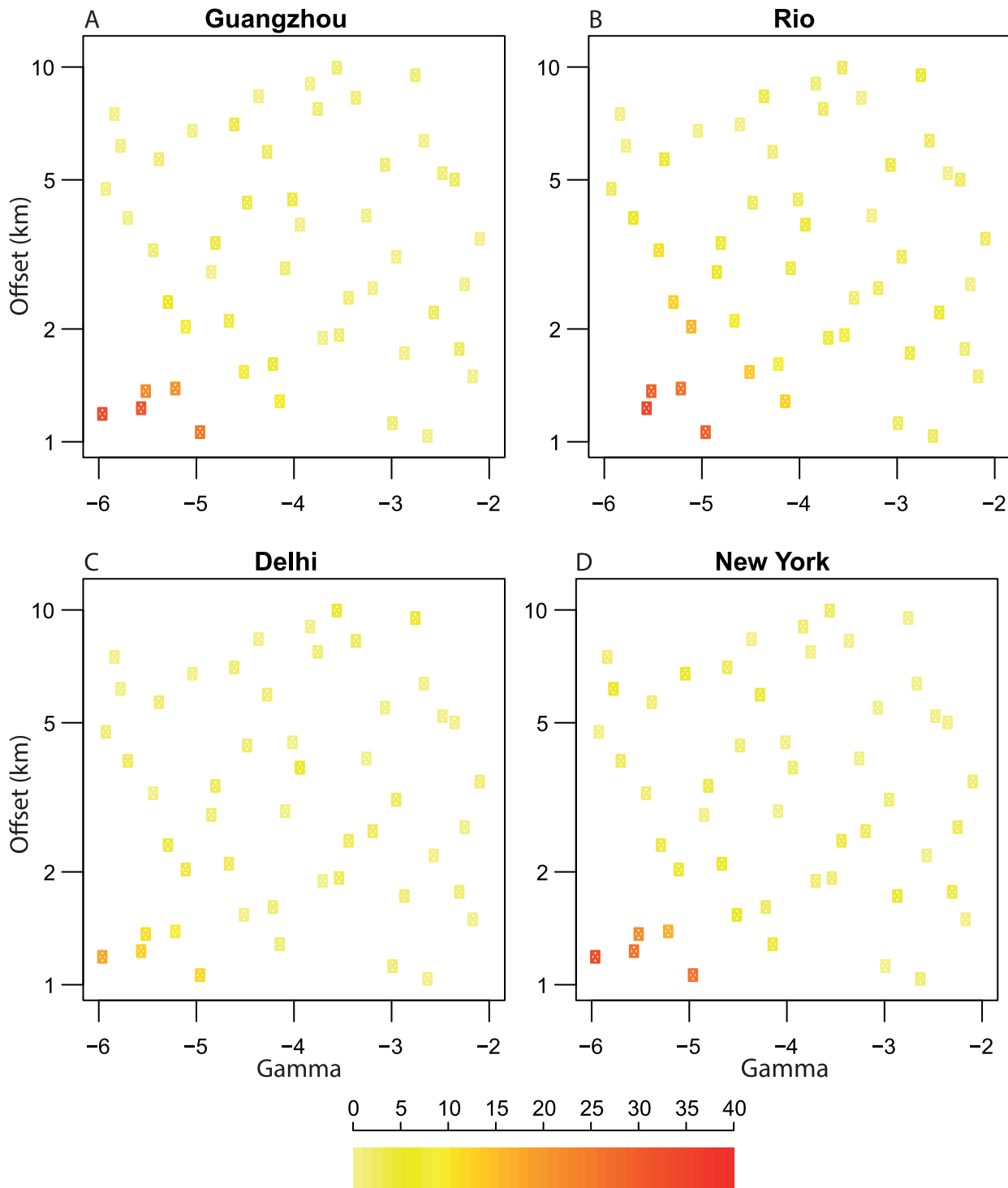
Figure 3. Latin Hypercube Sampling (LHS) results for four regions: Guangzhou, Rio, Delhi and New York. The LHS parameter sets varied γ and the offset (s_d) in the power law kernel, as described in the main text. We illustrate which parameter sets significantly affected the peak incidence at different resolutions by the colour of the point. The colour indicates the percentage change between the two points of lowest resolution and the two points of highest resolution, the colour-bar scale runs from 0% to 40%. Some parameter sets gave higher peak incidence at high resolution than low resolution (around −3% change), these were assumed to be caused by stochasticity as the number of runs was relatively small (10 repeats); the change for these was fixed at 0%. See Fig. S5 for plots of the trends in peak incidence for each parameter set.
doi:10.1371/journal.pcbi.1003561.g003

The rate at which susceptible individuals in pixel i became infected depended on (1) their risk of infection from those in pixel i, (2) the risk of infection from infected individuals in pixel j who travelled to i, (3) the risk of infection that susceptible individuals from i encountered when they travelled to j. Therefore, the *force of infection* or the average rate that susceptible individuals in pixel i
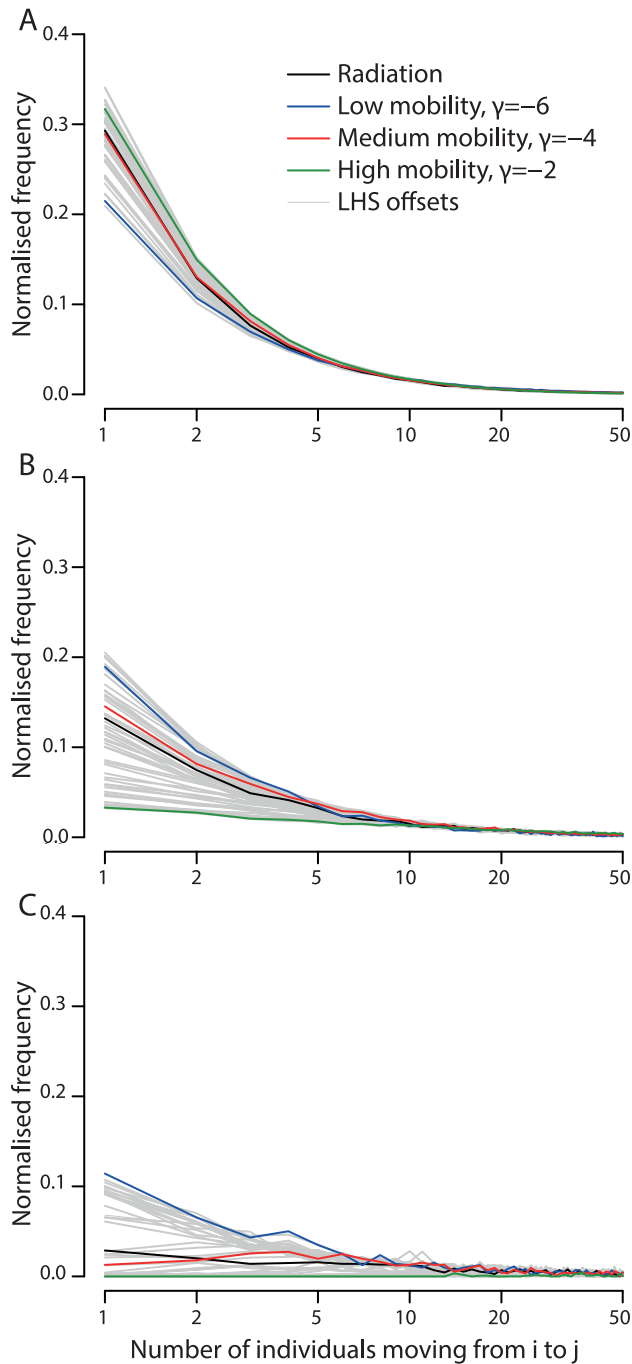
Figure 4. The number of individuals moving between populations (the flux) as defined by the offset gravity model (low, medium and high mobility and the LHS models) and for comparison the radiation model. We illustrate the mobility models at three resolutions: A: highest resolution, B: medium resolution and C: low resolution. The highest mobility model has higher average flux compared to the low and medium mobility models. At lower resolutions and high mobility large numbers of individuals move, but the distribution of the numbers moving is more uniform than at the highest resolution, causing the lines for the three mobility scenarios to swap order at different resolutions.
doi:10.1371/journal.pcbi.1003561.g004

became infected per time-step was:

$$\lambda_i = \beta \sum_{j=1}^{D} m_{ij} \frac{\sum_{l=1}^{D} m_{lj} I_l}{\sum_{p=1}^{D} m_{pj} N_p}. \tag{1}$$

where D was the total number of pixels and for any pixel j, N_j was the number of individuals, I_j was the number of infected individuals and infectious contacts were made with other individuals present in the pixel with rate β. Note that β is the same across all pixels; in future it may be of interest to vary the transmissibility across pixels (so β moves into the sum in Eqn. (1) as β_j).

The system of difference equations for a pixel i in the stochastic SIR model was (with the condition that all classes hold a whole number of individuals):

$$
\begin{aligned}
S_{(i,t+1)} &= S_{(i,t)} - N_{(i,S,t)}^{inf} \\
I_{(i,t+1)} &= I_{(i,t)} + N_{(i,S,t)}^{inf} - N_{(i,I,t)}^{rec} \\
R_{(i,t+1)} &= R_{(i,t)} + N_{(i,I,t)}^{rec},
\end{aligned}
\tag{2}
$$

where $N_{(i,v,t)}^{event}$ was the number of individuals in pixel i, state v (S, I or R) that experienced the event – infection or recovery – in time-step t. We ignored death in this model as we considered fairly short timescales and a non-fatal strain of influenza.

Each time-step t, the number of individuals experiencing each event ($N_{(i,v,t)}^{event}$) that occurred in pixel i and state v, with a population $N_{(i,v)}$ was determined in the following way:

1. For each pixel i, state v, the probability that any event would happen to an individual in that pixel and state was:

$$p_{(i,v)}^{any} = 1 - \exp(-(\sum_e H_e)\delta_t),$$

where the H_e were the rates for the events that may occur in that compartment, (e.g. the rate of recovery, recall that these parameters were chosen to reflect the natural history of influenza).

2. For each pixel i, state v, the total number of individuals who experienced an event $N_{(i,v,t)}^{any}$ was chosen from a Binomial distribution,

$$B(N_{(i,v)}, p_{(i,v)}^{any}).$$

3. The numbers of individuals experiencing each event ($N_{(i,v,t)}^{event}$) were drawn from a multinomial distribution with $N_{(i,v,t)}^{any}$ trials and the normalised selection probabilities $q_1, q_2, ..., q_j$ where

$$q_e = \frac{p_e}{(\sum_k p_k)}$$

and

$$p_e = 1 - \exp(-H_e \delta_t),$$

where p_e was the probability of each event occurring.

Mobility model

We used a mobility model to determine the relative frequency of potentially infectious contact. This was represented as a matrix m with entries m_{ij}, defined as the probability that for an individual from pixel i, given that the individual made a contact, this contact was with an individual from pixel j, so:

$$m_{ij} = N_j \kappa(r_{ij}) \left(\frac{1}{\sum_k N_k \kappa(r_{ik})} \right), \qquad (3)$$

where N_j was the total population in pixel j, $\kappa(r_{ij})$ was the interaction kernel defining the effect of the distance between pixels i and j on the contact between them. The kernel defines the relative probability of travelling between two pixels and not the absolute flux, similar to [15,17]. The factor $\frac{1}{\sum_k N_k \kappa(r_{ik})}$ normalised m and ensured that the rows sum to 1. The matrix m was used in the calculation of force of infection, Eqn. (1).

We used a variation of the offset power function for the kernel (similar to [9,15]):

$$\kappa(r_{ij}) = \frac{1}{1 + \left(s_d / r_{ij} \right)^\gamma} \qquad (4)$$

where s_d was the distance below which the kernel function saturated, we used $s_d = 2$km. The power γ determined the mixing between pixels, this was varied to give a range of mobilities but was always less than 0.

The next generation matrix and R_0

The next generation matrix, G, for the model with D pixels and force of infection λ_i (Eqn. (1)), can be defined (similar to [21]):

$$G_{ij} = N_i T \beta \sum_{k=1}^{D} \frac{m_{ik} m_{jk}}{\sum_{p=1}^{D} m_{pk} N_p} \qquad (5)$$

where T was the time spent infected (which depended on recovery rate α such that $T = 1/\alpha$, same for all pixels), N_i was the number of individuals in pixel i, infectious contacts were made with other individuals present in the pixel with rate β and m was the mobility matrix defined earlier. Then R_0 was equal to the spectral radius of this matrix $\rho(G)$ [36,37]. For this model, $\rho(G) = \beta T = R_0$, i.e. R_0 was independent of resolution and mobility; see Text S1 in Supporting Information S1 for full derivation.

Supporting Information

Figure S1 A simple example to illustrate how the resolution of the population is changed by aggregating multiple smaller units into larger units. The raster package in R is used to manipulate the population data. A: The population at its finest spatial resolution, 36 squares total. B: Four (2×2) of the smallest units from the population in A are combined into one pixel. C: Nine (3×3) of the smallest units (from A) are combined to make one pixel. D: The whole region is considered as one pixel. Note that the populations of the aggregated units are summed to find the total number of individuals in the new pixels.
(EPS)

Figure S2 The kernel function, $\kappa(r)$ against the distances, r, for the three different mobilities that we consider in the main results. The kernel is an offset power law function, $\frac{1}{1 + (s_d/r)^\gamma}$, where $s_d = 2$km and γ is -6, -4 and -2.
(EPS)

Figure S3 Details of the regions used in the main analysis. Histograms of the population densities of the four regions: Guangzhou, Rio, Delhi and New York and unprojected spatial maps of the populations. The central point is marked by a red dot and the area where the epidemic was seeded is marked in orange.
(EPS)

Figure S4 The kernels for the 50 parameter sets used in the Latin Hypercube Analysis of the model. The kernel is an offset power law function defined in the main text. The parameter sets vary γ between -6 and -2 and s_d between 1 and 10 km (on a log scale).
(EPS)

Figure S5 The impact of resolution on peak incidence for a range of parameter sets chosen by Latin Hypercube Sampling. The results are for four regions: Guangzhou, Rio, Delhi and New York. SLU are the Smallest Landscan Units (30″ by 30″ or approximately 1km²). The LHS parameter sets varied γ and the offset (s_d) in the power law kernel, as described in the main text. The colour of the line indicates the percentage change between the two points of lowest resolution and the two points of highest resolution, the colour-bar scale runs from 0% to 40%. Some parameter sets gave higher peak incidence at high resolution than low resolution (around -3% change), these were assumed to be caused by stochasticity as the number of runs was relatively small and the change was fixed at 0%.
(EPS)

Movie S1 The spatial spread of the epidemic in a theoretical population for the three mobility scenarios considered in the main text. From top to bottom is most restrictive ($\gamma = -6$) to least restrictive ($\gamma = -2$). On the left: the spread of an epidemic seeded in the centre of the region, indicated by the prevalence. On the right: peak incidence in the region. These plots are for the highest resolution of the theoretical region described in Fig. 1 of the main text.
(PDF)

Supporting Information S1 The spatial resolution of epidemic peaks. Text S1: Simplification of the next generation matrix. **Text S2:** Final epidemic size.
(PDF)

Table S1 A review of current data and studies on mobility patterns in humans. There is a lack of empirical data detailing why people travel, mode and distance travelled, divided by age and gender.
(PDF)

Table S2 Details of the 4 regions used in the main paper. All regions are 49×29 cells in size (2401 cells total), area varies according to latitude and longitude. Rio has a large number of zero regions because it is on the coast. Delhi has the highest

population density but also the highest variance in population sizes. The map in Fig. 2 indicates the locations of these regions on the world map. Fig. S3 contains histograms of the population densities and spatial maps.
(PDF)

Author Contributions

Conceived and designed the experiments: HLM SR. Performed the experiments: HLM. Analyzed the data: HLM SR. Wrote the paper: HLM SR.

References

1. Anderson RM, Fraser C, Ghani AC, Donnelly CA, Riley S, et al. (2004) Epidemiology, transmission dynamics and control of SARS: the 2002–2003 epidemic. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 359: 1091–1105.
2. Sattenspiel L, Herring DA (2003) Simulating the effect of quarantine on the spread of the 1918–19 flu in central Canada. Bulletin of mathematical biology 65: 1–26.
3. Chowell G, Ammon C, Hengartner N, Hyman J, et al. (2006) Transmission dynamics of the great influenza pandemic of 1918 in Geneva, Switzerland: Assessing the effects of hypothetical interventions. Journal of Theoretical Biology 241: 193–204.
4. Lipsitch M, Riley S, Cauchemez S, Ghani AC, Ferguson NM (2009) Managing and reducing uncertainty in an emerging influenza pandemic. New England Journal of Medicine 361: 112–115.
5. Yu H, Cowling BJ, Feng L, Lau EH, Liao Q, et al. (2013) Human infection with avian influenza A H7N9 virus: an assessment of clinical severity. The Lancet 382: 138–145.
6. Gao R, Cao B, Hu Y, Feng Z, Wang D, et al. (2013) Human infection with a novel avian-origin influenza A (H7N9) virus. New England Journal of Medicine 368: 1888–1897.
7. Cauchemez S, Van Kerkhove M, Riley S, Donnelly C, Fraser C, et al. (2013) Transmission scenarios for Middle East Respiratory Syndrome Coronavirus (MERS-CoV) and how to tell them apart. Euro Surveill 18: 20503.
8. Breban R, Riou J, Fontanet A (2013) Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. The Lancet 382: 694–9.
9. Truscott J, Ferguson N (2012) Evaluating the adequacy of gravity models as a description of human mobility for epidemic modelling. PLoS Computational Biology 8: e1002699.
10. Mao L, Bian L (2010) Spatial–temporal transmission of influenza and its health risks in an urbanized area. Computers, Environment and Urban Systems 34: 204–215.
11. Tsai M, Chern T, Chuang J, Hsueh C, Kuo H, et al. (2010) Efficient simulation of the spatial transmission dynamics of influenza. PloS one 5: e13292.
12. Finkenstädt B, Grenfell B (1998) Empirical determinants of measles metapopulation dynamics in England and Wales. Proceedings of the Royal Society of London Series B: Biological Sciences 265: 211–220.
13. Xia Y, Bjrnstad ON, Grenfell BT (2004) Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics. The American Naturalist 164: 267–281.
14. Bolker B, Grenfell B, Bolker B, Grenfell B (1995) Space, persistence and dynamics of measles epidemics. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 348: 309–320.
15. Riley S, Ferguson N (2006) Smallpox transmission and control: spatial dynamics in Great Britain. Proceedings of the National Academy of Sciences 103: 12637–12642.
16. Hall I, Egan J, Barrass I, Gani R, Leach S (2007) Comparison of smallpox outbreak control strategies using a spatial metapopulation model. Epidemiology and infection 135: 1133–1144.
17. Ferguson NM, Cummings DA, Cauchemez S, Fraser C, Riley S, et al. (2005) Strategies for containing an emerging influenza pandemic in Southeast Asia. Nature 437: 209–214.
18. Wu JT, Riley S, Fraser C, Leung GM (2006) Reducing the impact of the next influenza pandemic using household-based public health interventions. PLoS Medicine 3: e361.
19. House T, Keeling MJ (2008) Deterministic epidemic models with explicit household structure. Mathematical biosciences 213: 29–39.
20. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, et al. (2010) Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. Journal of computational science 1: 132–145.
21. Wu J, Riley S, Leung G (2007) Spatial considerations for the allocation of prepandemic influenza vaccination in the United States. Proceedings of the Royal Society B: Biological Sciences 274: 2811–2817.
22. Department for Transport (2011). National travel survey. Available: https://www.gov.uk/government/organisations/department-for-transport/series/national-travel-surv
23. Arino J, Van den Driessche P (2003) A multi-city epidemic model. Mathematical Population Studies 10: 175–193.
24. Ma J, Earn D (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease. Bulletin of Mathematical Biology 68: 679–702.
25. Andreasen V (2011) The final size of an epidemic and its relation to the basic reproduction number. Bulletin of Mathematical Biology 73: 2305–2321.
26. Oakridge National Laboratory. Landscan global population data. Available: http://www.ornl.gov/sci/landscan/.
27. Iman RL (2008) Latin hypercube sampling. Wiley Online Library.
28. Simini F, González MC, Maritan A, Barabási AL (2012) A universal model for mobility and migration patterns. Nature 484: 96–100.
29. Mylius SD, Hagenaars TJ, Lugnér AK, Wallinga J (2008) Optimal allocation of pandemic influenza vaccine depends on age, risk and timing. Vaccine 26: 3742–3749.
30. Klepac P, Bjørnstad ON, Metcalf CJE, Grenfell BT (2012) Optimizing reactive responses to outbreaks of immunizing infections: balancing case management and vaccination. PloS one 7: e41428.
31. Poland GA (2006) Vaccines against avian influenza - a race against time. New England Journal of Medicine 354: 1411–1413.
32. Hessel L (2009) Pandemic influenza vaccines: meeting the supply, distribution and deployment challenges. Influenza and other respiratory viruses 3: 165–170.
33. Mossong J, Hens N, Jit M, Beutels P, Auranen K, et al. (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 5: e74.
34. Eames KT, Tilston NL, Edmunds WJ (2011) The impact of school holidays on the social mixing patterns of school children. Epidemics 3: 103–108.
35. Merler S, Ajelli M, Pugliese A, Ferguson NM (2011) Determinants of the spatiotemporal dynamics of the 2009 H1N1 pandemic in Europe: implications for real-time modelling. PLoS computational biology 7: e1002205.
36. Diekmann O, Heesterbeek J, Metz J (1990) On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology 28: 365–382.
37. van den Driessche P, Watmough J (2002) Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences 180: 29–48.