Research

Author for correspondence:
Leon Danon
e-mail: l.danon@bristol.ac.uk

A spatial model of COVID-19 transmission in England and Wales: early spread, peak timing and the impact of seasonality

Leon Danon[1], Ellen Brooks-Pollock[2,3], Mick Bailey[2] and Matt Keeling[4]

[1]Department of Engineering Mathematics, [2]Bristol Veterinary School, and [3]NIHR Health Protection Research Unit (HPRU) in Behavioural Science and Evaluation, Population Health Sciences, University of Bristol, Bristol BS8 1QU, UK
[4]Mathematics Institute, and School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

LD, 0000-0002-7076-1871; EB-P, 0000-0002-5984-4932; MK, 0000-0003-4639-4765

An outbreak of a novel coronavirus was first reported in China on 31 December 2019. As of 9 February 2020, cases have been reported in 25 countries, including probable human-to-human transmission in England. We adapted an existing national-scale metapopulation model to capture the spread of COVID-19 in England and Wales. We used 2011 census data to inform population sizes and movements, together with parameter estimates from the outbreak in China. We predict that the epidemic will peak 126 to 147 days (approx. 4 months) after the start of person-to-person transmission in the absence of controls. Assuming biological parameters remain unchanged and transmission persists from February, we expect the peak to occur in June. Starting location and model stochasticity have a minimal impact on peak timing. However, realistic parameter uncertainty leads to peak time estimates ranging from 78 to 241 days following sustained transmission. Seasonal changes in transmission rate can substantially impact the timing and size of the epidemic. We provide initial estimates of the epidemic potential of COVID-19. These results can be refined with more precise parameters. Seasonal changes in transmission could shift the timing of the peak into winter, with important implications for healthcare capacity planning.

This article is part of the theme issue 'Modelling that shaped the early COVID-19 pandemic response in the UK.

1. Introduction

An outbreak of a novel coronavirus, recently renamed COVID-19, was first reported from Wuhan, China on 31 December 2019. During January 2020, the outbreak spread to multiple cities in China, and the first cases started appearing outside China. By the end of January 2020, 9720 cases had been confirmed in China, with 106 confirmed cases outside China across 19 different countries [1].

Epidemiological analysis of the outbreak was quickly used to start estimating the most relevant parameters, such as the basic reproduction number, the serial interval, the incubation period and the case fatality rate [2–7]. Initial estimates suggested that the reproduction number was between 2 and 3 and the case fatality rate was less than 4% [8]. Control of spread by contact tracing and isolation appears to be challenging, given what is currently known about the virus [9].

Mathematical models are useful tools for understanding and predicting the possible course of an outbreak, given a set of underlying assumptions. Here, we adapt a metapopulation model of disease transmission in England and Wales to capture the spread of COVID-19 [10]. The aim is to provide predictions about the likely timing of the peak of the epidemic in England and Wales and spatial features of spread.
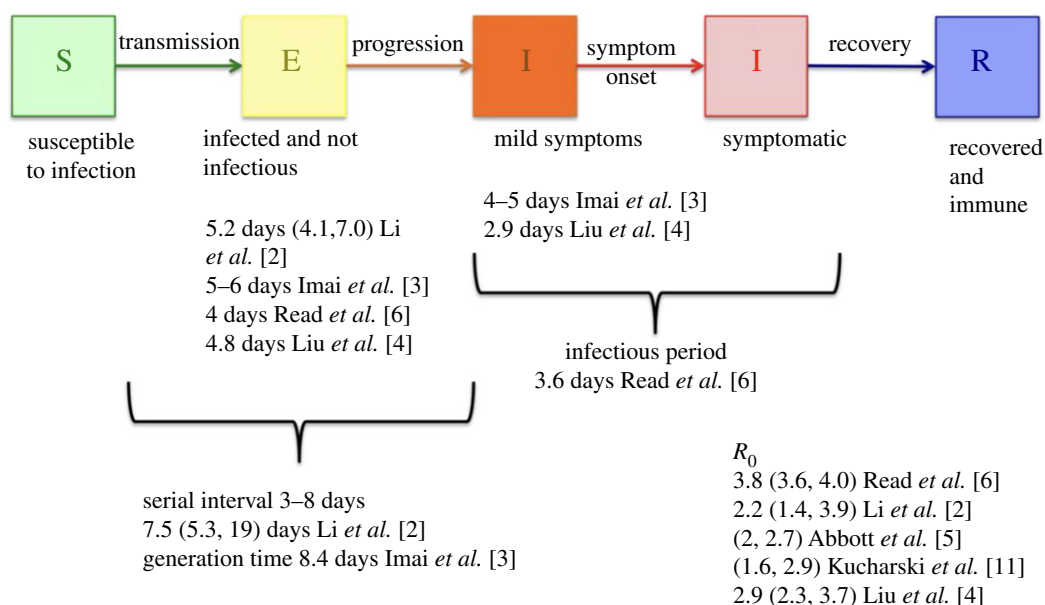
Figure 1. Model structure within each ward, together with associated parameters estimated from the literature. (Online version in colour.)

2. Methods

(a) Model description

We use an existing national-scale stochastic metapopulation model of disease transmission in England and Wales. The model structure is based on the metapopulation model described in detail in Danon *et al.* [10]. In this model, the population is divided into electoral wards. Because of the changes in data linkage, we restricted the model to England and Wales, whereas the original model covered Great Britain.

(b) Movement between wards

Transmission between wards occurs via the daily movement of individuals. For each ward, we assume that individuals contribute to the force of infection in their 'home' ward during the night and their 'work' ward during the day. Regular movements that model commuting behaviour are included in the model as well as irregular movements that represent the population that does not commute to work. See Danon *et al.* [10] for further details.

(c) Population and movement data

Data for population and movement of individuals come from the 2011 census of the United Kingdom. The population size of each of the 8570 electoral wards is available directly from the Office of National Statistics (ONS) website. The number of individuals moving between locations is also available from the ONS website, but at the level of census output areas (OAs). We aggregated the data from OA level to electoral wards level. The spatial location of electoral ward centres is extracted from maps available from the ONS websites.

(d) COVID-19 specific parameters

We use a Susceptible-Exposed-Infectious-Infectious-Recovered (SEIIR) model within each ward to capture the progression of disease within an individual (figure 1). Initial analyses used SARS-like parameters for the incubation period and infectious period, which now appear to differ from COVID-19 [4,11]. Li *et al.* [2] analysed data on 425 cases reported in Wuhan in China and fitted a lognormal distribution to the incubation period, and a gamma distribution for the serial interval. The infectious period for SARS was estimated as the serial interval minus the incubation period, but as Li *et al.* did not report the correlation between incubation period and serial interval, we

Table 1. Biological parameters and distributions used in the model.

parameter	values and distribution	reference
incubation period	lognormal (meanlog = 5.2, s.d.log = 0.35)	Li *et al.* [2]
reproduction number	gamma (scale = 2.2/100, shape = 100)	Li *et al.* [2]
infectious period	uniform (2,3)	estimated from the mean serial interval (7.5 days) minus the mean incubation period (5.2 days) from Li *et al.* [2]

were not able to estimate the infectious period distribution from the data but used a uniform distribution between 2 and 3 days, to give a mean serial interval of approximately 7–8 days, in line with current estimates. We used two infectious states to represent a mildly symptomatic or prodromal period and a period with more pronounced symptoms. In the absence of data on the relative magnitude of these two infections states, we assumed the same length of time in each infectious state and assumed that each state was equally infectious. We sampled from each of the distributions 100 times independently (table 1).

(e) Initialization and baseline model

The census data are used to initialize the population sizes within each of the 8570 wards that hold between 200 and 10 000 individuals. At the start of the model, all individuals are assumed to be susceptible to infection with no underlying immunity in the population. To seed infection in a ward, we move five individuals (non-commuters) from the susceptible compartment to the first infectious state.

We investigated a range of starting scenarios by seeding the infection in example wards in London, Birmingham, Liverpool, Bristol, Manchester, Sheffield and Cardiff. We also investigated a generalized epidemic case, where cases were simultaneously imported in three different locations, seeding the infection in London, Birmingham and Manchester simultaneously on day 1.
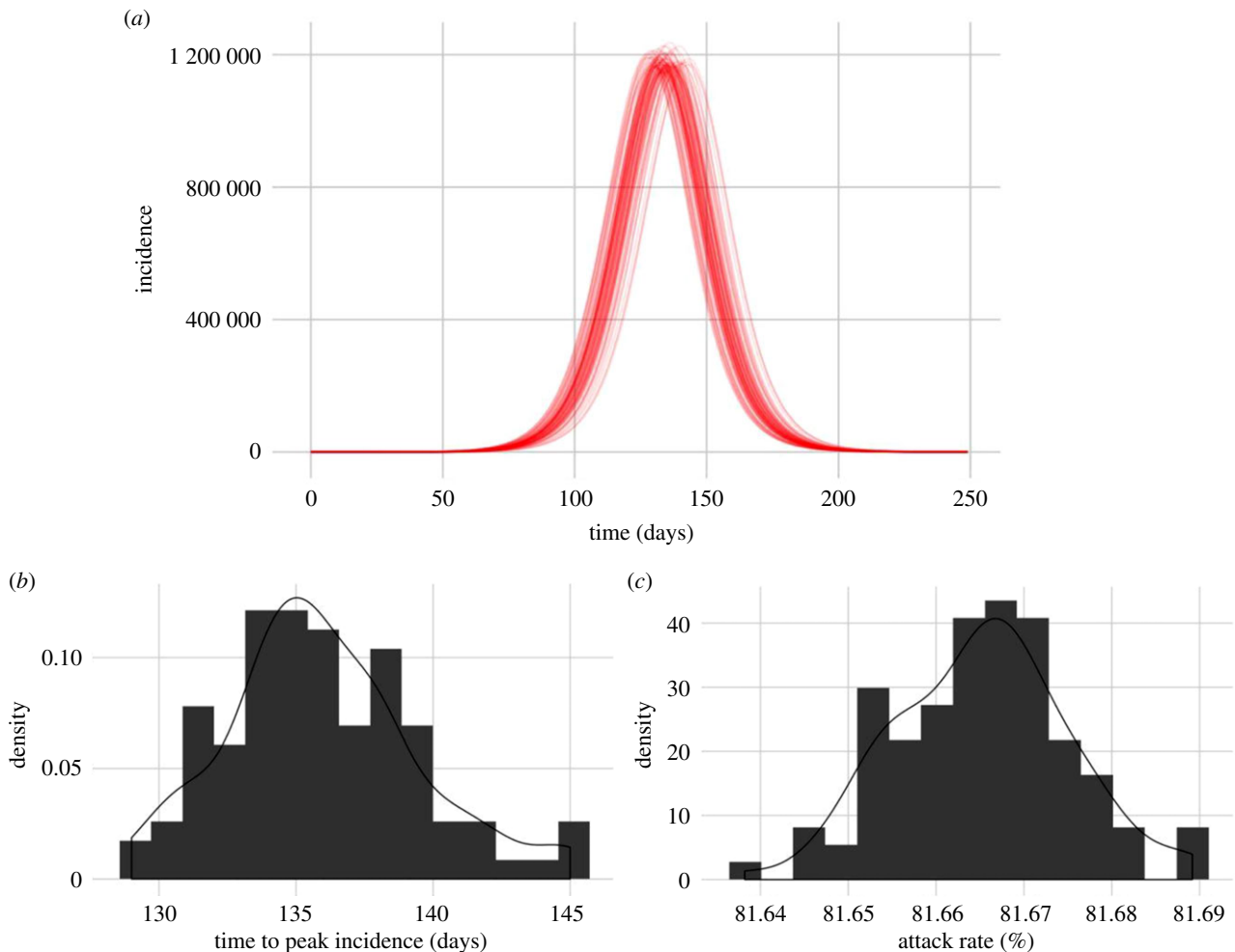
Figure 2. The number of cases of COVID-19 in England and Wales in the absence of any control measures, 100 realizations of the spatial model, seeded in Brighton, using best-guess parameters from Li et al. [2] (a) Daily infection dynamics. (b) The distribution of predicted time to peak incidence. (c) The distribution of predicted attack rate. (Online version in colour.)

(f) Impact of seasonality

We investigated the impact of a seasonally affected transmission rate, to capture potential decreased transmission during the summer months. We captured seasonal transmission by replacing the constant transmission rate with a time-varying transmission rate given by:

$$\text{transmission rate} = \beta\left(1 - \frac{m}{2}\left(1 - \cos\frac{2\pi t}{365}\right)\right),$$

where m is the magnitude of the seasonal difference in transmission, ranging from $m = 0$ (no seasonality) to $m = 1$ (maximum seasonality with no transmission at the peak of the summer).

(g) Epidemic characteristics

From the model, we extracted the total number of infections per day, as the number of individuals in both of the Infectious states, and the number of infected wards per day as the total number of wards with at least one individual in one of the two Infectious states. The spatial growth of the epidemic in England and Wales was visualized using interactive maps. We estimated the timing of the epidemic peak from the aggregated epidemic curve and calculated 95% prediction intervals from the model simulations.

(h) Implementation and data availability

The model is coded in C and is available on GitHub (http://github.com/ldanon/MetaWards), with an updated implementation in python (http://metwards.org). The data for

parameterizing the model are freely available from the ONS website or can be downloaded with the code at the GitHub repository.

3. Results

We predict that, in the absence of any interventions, a disease with 'best-guess' COVID-19-like parameters will peak at a median of 133 days (range 126–147 days) following the start of person-to-person transmission in England and Wales. Intrinsic model stochasticity is responsible for variation between model runs. Using exactly the same parameters and seeding the infection in the same initial wards resulted in a difference in peak timing of $+/-10$ days (figure 2). The attack rate for best-guess parameters had a median of 45 799 874 (81.67% range 81.64–81.69), with a peak incidence median of 1 116 692.

Model predictions are highly sensitive to parameter values and incorporating parameter uncertainty increases model variability substantially. In the absence of any control measures, all predictions resulted in epidemics that peaked within a year from the start of person-to-person transmission in England and Wales. Estimates of peak time ranged from 78 days to 241 days, albeit with a low probability (figure 3). The model peak time was particularly sensitive to the value of incubation period and the transmission rate; these were chosen from ranges given in table 1.

The initial location of cases had some, but limited impact on the timing of the epidemic in England and Wales.
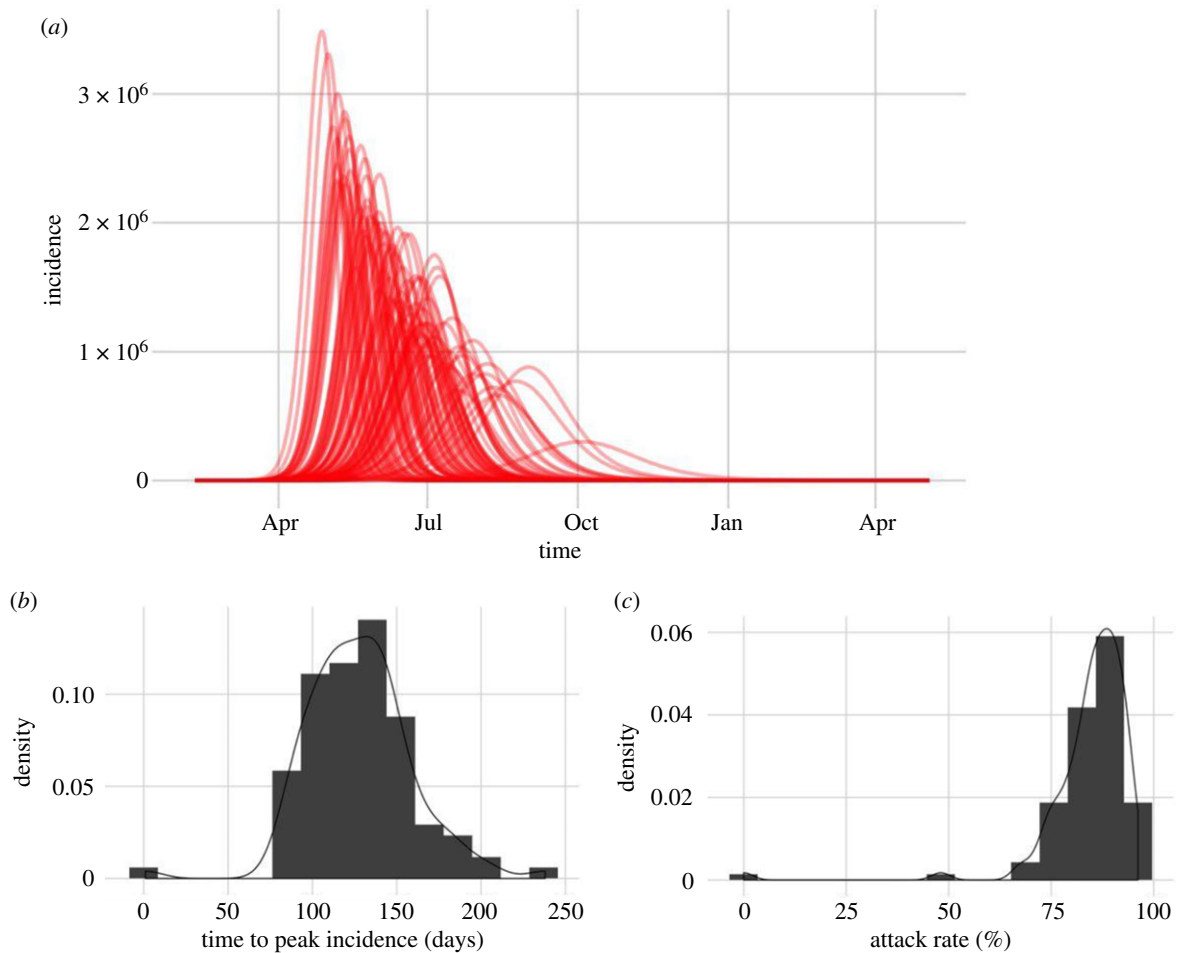
Figure 3. The variability in predicted epidemic curves for a COVID-19 outbreak in England and Wales, seeded in Brighton, in the absence of any control measures. Unlike in figure 2, here we incorporate measured parameter uncertainty. (*a*) Daily infection dynamics. (*b*) The distribution of predicted time to peak incidence. (*c*) The distribution of predicted attack rate.
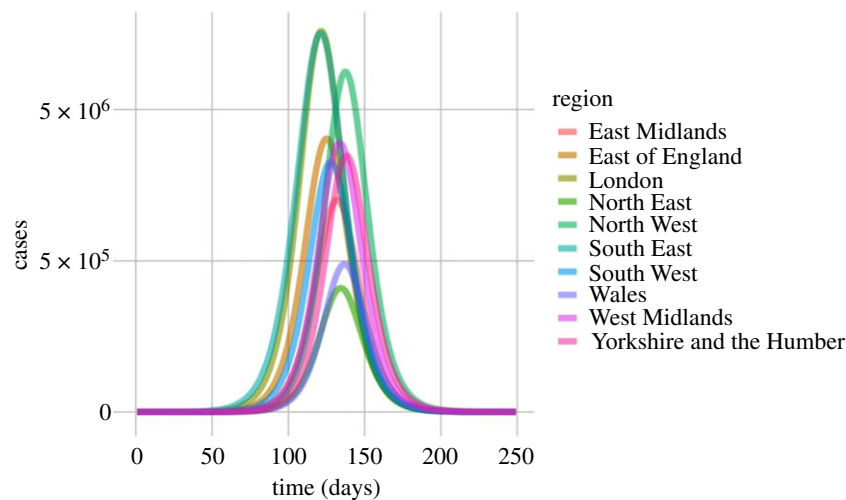

Figure 4. Predicted epidemic curves for a COVID-19 outbreak broken down by region for England and Wales. (Online version in colour.)

Epidemics seeded in Brighton, London, Birmingham and Sheffield resulted in synchronized epidemics in England, reaching urban areas first followed by rural areas. Epidemics started in Cardiff had a slower time to peak but still resulted in a generalized outbreak. We also investigated a generalized seeding scenario, where cases were simultaneously imported in three different locations, seeding the infection in London, Birmingham and Manchester at the same time.

Spatially, some disaggregation between England and Wales regions is observed. An outbreak starting in Brighton, (South

East England) peaks in London and the South East first, and North East England, Yorkshire and Humber and Wales last, with a 10-day lag between regional peaks (figure 4).

Figure 5 shows the impact of seeding location on the spatial distribution of cases in more fine-grained detail. Seeding the infection in a single city leads to an earlier peak burden in that city. Setting London and Birmingham as seeds led to the most synchronized countrywide outbreaks. Seeding in London resulted in other major cities peaking two weeks after peak burden in London. By contrast, seeding
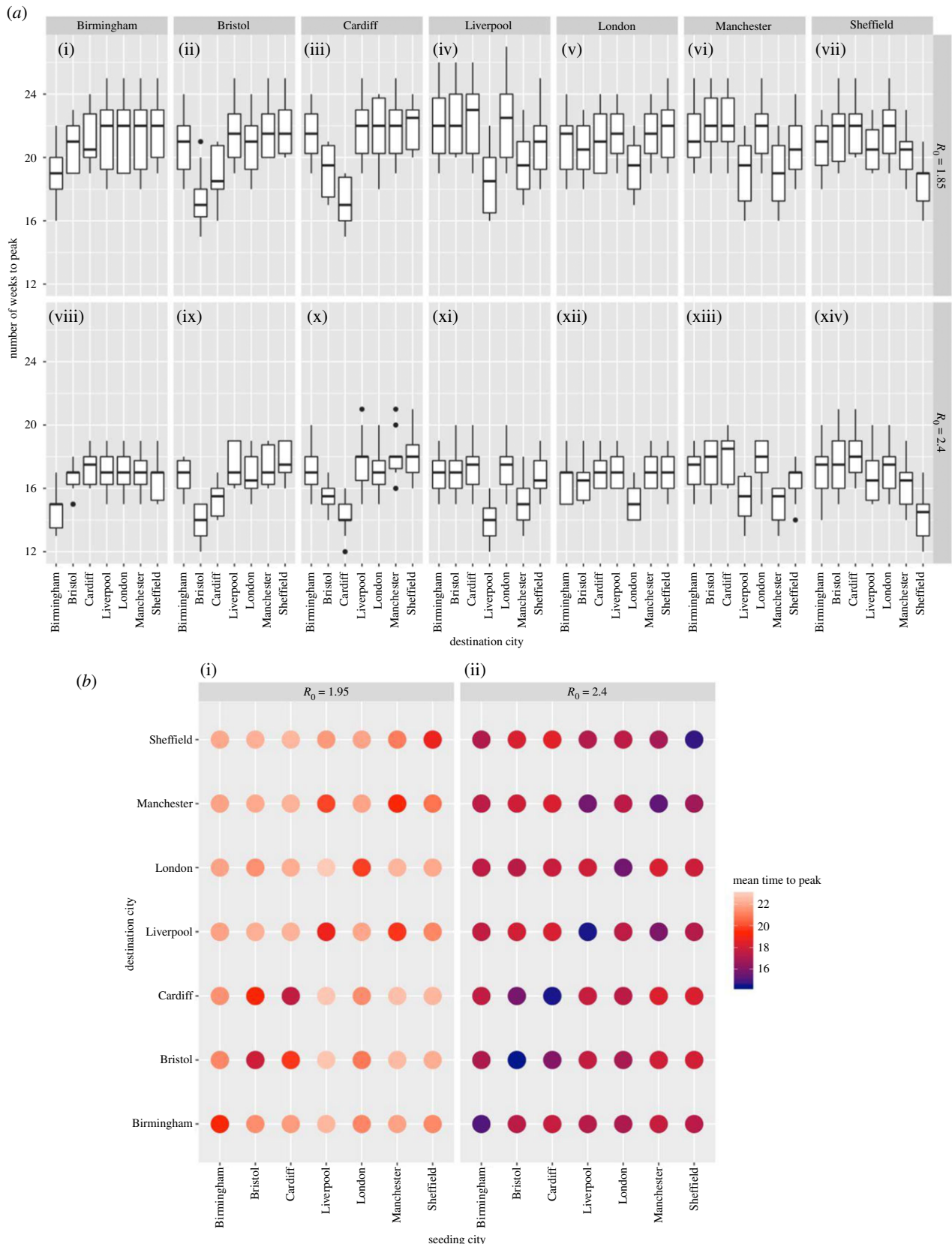
Figure 5. (*a*) Peak time in major cities from various starting locations. Each panel is a starting location and the box plots show the distribution in peak times in each destination city from 10 runs. (*b*) Average peak time in each city shown as a matrix from the start location. (Online version in colour.)

in Birmingham led to other major cities peaking four weeks after peaking in Birmingham.

The spatial proximity of cities can be observed reflected in the time to peak. For instance, when seeding the infection in Cardiff, a peak is observed in Bristol two weeks later, with other cities peaking 4 to 5 weeks after seeding. No epidemic

scenarios we considered had a temporal spread between major cities of more than eight weeks (figure 5).

Epidemics resulting from multiple importations were modelled by seeding at multiple locations simultaneously. Seeding the infection in London, Birmingham and Manchester simultaneously results in early peaks in those cities,
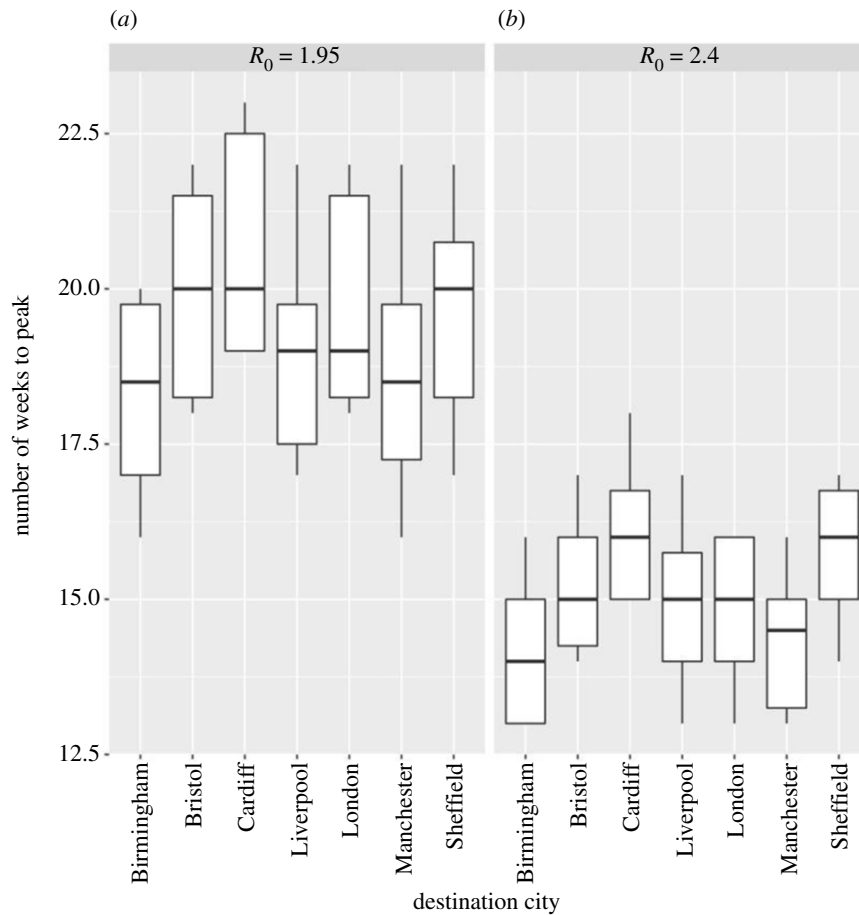
Figure 6. Peak timing in major cities for a generalized epidemic with multiple initial seed locations. Box plots represent the variability between 10 parameter sets with the same R_0 (1.95, 2.4) and mean doubling time (6.6, 4.7 days).
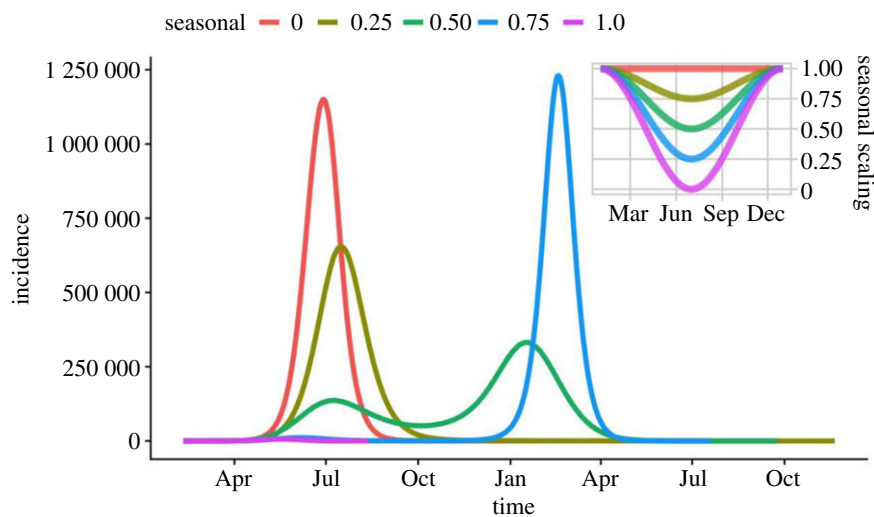

Figure 7. Effect of seasonal changes in transmission rate, assuming a reduction in transmission over the summer. (Main panel) Incidence over time, for different values of seasonal scaling. (Inset) Variation of scaling term for the course of one year, with transmission being at its lowest in July. A strong suppression in the initial growth phase may generate a perverse outcome of a second peak in the winter. (Online version in colour.)

followed closely by Liverpool, Bristol, Sheffield and Cardiff. This scenario leads to synchronous epidemics with small differences between peak times in major cities and peaks appearing earlier in seeding locations (figure 6).

However, seasonality in transmission has a large impact on epidemic timing, peak incidence and final attack rates. Assuming no difference in transmission rate during the year leads to a single large epidemic peak after approximately four months (June if transmission starts in February), as above. With a

25% reduction in transmission the epidemic is smaller and peaks later, reducing the overall attack rate by 20%. A 50% reduction in transmission results in a smaller epidemic before the summer, followed by a resurgence in cases in the following winter. The attack rate is 10% less than a non-seasonal epidemic. A 75% reduction in transmission over the summer resulted in a delayed large outbreak, but with a similar attack rate. If transmission decreases to zero over the summer, then the resulting outbreak experiences stochastic

Table 2. Effect of seasonal variation on the timing (shown in days following initial seeding), the height of the peak and the final attack rate.

seasonal term	timing of peak	incidence at peak	final attack rate
0	139	1 172 819	81.9
0.25	159	615 599	65.0
0.50	343	330 311	69.4
0.75	375	1 227 280	80.3
1.0	100	6547	0.53

fadeout, the peak is dramatically reduced, with a final attack rate of less than 1% due to extinction (figure 7 and table 2). This scenario is unlikely as the reintroduction of infection from outside England and Wales would likely lead to further waves of infection.

4. Discussion

We predict that, in the absence of control measures and with no seasonality in transmission, the introduction of COVID-19 in England and Wales has the potential to result in a synchronized outbreak that peaks at around four months following the start of person-to-person transmission. Our findings suggest that the height of the epidemic and the attack rate is highly dependent on the seasonality of transmission and that even small changes in transmission risk can lead to large changes in attack rate due to the spatial disaggregation of the population at risk.

A combination of control measures and seasonal changes in transmission rate could shift the peak of the outbreak to the winter of 2020/21, with little effect on the final attack rate. If contact tracing and isolation efforts succeed in reducing transmission but are unable to control the epidemic [9], an additional influx of severe COVID-19 cases may exacerbate existing challenges with winter healthcare demand. A careful analysis of the impact of control measures on the timing of incidence of severe cases is warranted.

The strength of this model lies in the spatial heterogeneity which tempers transmission. As a comparison, an equivalent non-spatial model results in the epidemic peaking after 34 days, nearly four times faster than this spatial model, and would be unable to capture the interaction between spatial transmission and seasonality. The estimated total number of people infected in the spatial model is marginally smaller than for a non-spatial model, as the infection has the opportunity to die-out in local parts of the country. As the model framework was developed and published in 2009, it was possible to re-deploy the model for these new circumstances; developing such a model from scratch during an outbreak would be a significant challenge.

A key element missing from our model is morbidity, mortality and the treatment of cases. The model in its current form predicts the total number of infections in the community rather than diagnosed cases. Observations from China suggest that many cases have mild symptoms and that only around 5% of cases have been reported and diagnosed [3]. The parameter estimates we used from China appear to be substantially different to previous coronaviruses [6]. Should COVID-19 continue spreading the UK it will become possible to get UK-specific parameter estimates and improve prediction accuracy.

As with all modelling, it is impossible to capture the full complexity of an epidemic. In this model, the major assumptions are that we have assumed that there is no change in behaviour during the course of the epidemic. In practice, as the epidemic starts spreading in England and Wales, there may well be a systematic change in behaviour as was seen during the H1N1 influenza pandemic in 2009. We have not included any age effects, such as differential mixing, susceptibility or infectiousness. That means that we are not able to investigate the impact of school closures or the impact of the summer holidays, which had a large impact on the H1N1 influenza pandemic in 2009.

References

1. World Health Organization (WHO). 2020 *Novel Coronavirus (COVID19) situation report-23.* Geneva, Switzerland: WHO.

2. Li Q *et al.* 2020 Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *N. Eng. J. Med.* **382**, 1199–1207. (doi:10.1056/NEJMoa2001316)

3. Imai N, Cori A, Dorigatti I, Baguelin M, Donnelly CA, Riley S, Ferguson NM. 2020 *Report 3: transmissibility of 2019-nCoV.* London, UK: Imperial College London.

4. Liu T *et al.* 2020 Transmission dynamics of 2019 novel coronavirus (2019-nCoV). *bioRxiv* 2020.01.25.919787. (doi:10.1101/2020.01.25.919787)

5. Abbott S, Hellewell J, Munday J, Funk S, Funk S. 2020 The transmissibility of novel Coronavirus in the early stages of the 2019–20 outbreak in Wuhan: exploring initial point-source exposure sizes and durations using scenario analysis.

Wellcome Open Res. **5**, 17. (doi:10.12688/wellcomeopenres.15718.1)

6. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. 2020 Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* 2020.01.23.20018549-2020.01.23.20018549. (doi:10.1101/2020.01.23.20018549)

7. Wu JT, Leung K, Leung GM. 2020 Nowcasting and forecasting the potential domestic and

international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697. (doi:10.1016/S0140-6736(20)30260-9)

8. Dorigatti I *et al*. 2020 *Report 4: severity of 2019-novel coronavirus (nCoV)*. London, UK: Imperial College London.

9. Hellewell J *et al*. 2020 Feasibility of controlling 2019-nCoV outbreaks by isolation of cases and contacts. *medRxiv* 2020.02.08.20021162. (doi:10.1101/2020.02.08.20021162)

10. Danon L, House T, Keeling MJ. 2009 The role of routine versus random movements on the spread of

disease in Great Britain. *Epidemics* **1**, 250–258. (doi:10.1016/j.epidem.2009.11.002)

11. Kucharski AJ *et al*. 2020 Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 553–558. (doi:10.1016/S1473-3099(20)30144-4)