A COVID-19 Model for Local Authorities of the United Kingdom

Swapnil Mishra[1,*], James A. Scott[2,*], Daniel J. Laydon[1], Harrison Zhu[2], Neil M. Ferguson[2], Samir Bhatt[1], Seth Flaxman[2], Axel Gandy[2,*]

[1]*MRC Centre for Global Infectious Disease Analysis, Abdul Latif Jameel Institute for Disease and Emergency Analytics (J-IDEA), Imperial College London, London, UK*

[2] *Department of Mathematics, Imperial College London, London, UK*

* Contributed equally

E-Mail: a.gandy@imperial.ac.uk

Summary. We propose a new framework to model the COVID-19 epidemic of the United Kingdom at the local authority level. The model fits within a general framework for semi-mechanistic Bayesian models of the epidemic based on renewal equations, with some important innovations, including a random walk modelling the reproduction number, incorporating information from different sources, including surveys to estimate the time-varying proportion of infections that lead to reported cases or deaths, and modelling the underlying infections as latent random variables. The model is designed to be updated daily using publicly available data. We envisage the model to be useful for now-casting and short-term projections of the epidemic as well as estimating historical trends. The model fits are available on a public website, `https://imperialcollegelondon.github.io/covid19local`. The model is currently being used by the Scottish government to inform their interventions.

1. Introduction

Surveillance systems are vital to combat the spread of the SARS-CoV-2 epidemic. In the UK, publicly available estimates of current infections and the reproduction number (i.e. the number of secondary infections per infection) are primarily available at national and regional level (9 geographic regions of England) (Department of Health and Social Care , 2020). However, with an evolving epidemic localised trends at subnational level are very important. To identify trends and facilitate monitoring at a local level, we implement a semi-mechanistic Bayesian transmission model for SARS-CoV-2 at the local authority (LA) level for the UK. The model assesses and projects the evolution of the epidemic and estimates the time-varying reproduction number for local areas. We apply our analysis to the UK, although our approach is applicable to any country where local data on cases and deaths are available.

We have extended the Flaxman et al. (2020) model for each local authority by incorporating four innovations. First, the model incorporates reported cases in addition to deaths. Second, survey data from the (Office for National Statistics, 2020a) and from the Real-time Assessment of Community Transmission study (REACT Study, 2020) is used to calibrate estimates of the (unobserved) number of true infections. Third, the

model incorporates a time-varying infection fatality rate (IFR), the fraction of infections that leads to deaths, and a time-varying infection ascertainment rate (IAR), the fraction of infections identified as positive cases. Fourth, infections are modelled as a random process, and not merely a deterministic function of R_t, previous infections, and a given serial-interval (generation distribution), which better accounts for variability in areas with low infection numbers.

Regularly updated results from the model are presented at covid19local.† The model is currently being used by the Scottish government in their response to their epidemic (Scottish Government, 2020, issue 24 to now).

2. Data

We combine data from national statistics and public health bodies across the UK. Reported cases for England are taken from the UK Coronavirus dashboard (Public Health England, 2020), for Wales from (Public Health Wales, 2020). For England and Wales, information about deaths are from (Office for National Statistics, 2020b). Cases and deaths for Northern Ireland are from (Department of Health, Northern Ireland, 2020) and for Scotland are from (Scottish Health and Social Care Open Data, 2020).

For reported cases, the model uses the date of specimen collection. To account for reporting variations within a given week, we aggregate daily case and death data by week. We omit the last three days of data while fitting our model to account for reporting delays. ONS and REACT survey data are used to calibrate estimates of infections produced by our model.

Our model works at the local authority level - in England, where there are upper and lower tier local authorities, we work with lower tier local authorities (LTLAs). The other nations of the UK do not have this subdivision. In total we fit to 391 different areas, that consists of the UK, nations in the UK, geographic regions of England and all local areas in the UK.

3. Model

Flaxman et al. (2020) introduced a Bayesian semi-mechanistic framework for estimating the transmission intensity of SARS-CoV-2. The model is based on the renewal equation (Mishra et al., 2020), and uses R_t to generate new infections. We modify this framework for use with local authorities. In this section, we outline the model; more details are in Section S1 of the supplementary materials. Section 6 discusses the reasons for our modelling choices.

Let i_t be the number of infections on a given day t in a given area. The basic model in (Flaxman et al., 2020) uses the following renewal equation:

$$i_t = R_t \sum_{\tau=0}^{t-1} i_\tau g_{t-\tau}, \tag{1}$$

where R_t is the real-time reproduction number (i.e. the number of secondary infections per infection), and g_k, $k = 1, \ldots$ is the generation distribution, i.e. a probability mass

†https://imperialcollegelondon.github.io/covid19local

function determining the time between two infections. R_t is random and can be flexibly modelled.

Eq (1) can be modified to allow infections to be a latent process and account for population effects (decreasing susceptible population over time) as follows (Scott et al., 2020; Bhatt et al., 2020):

$$i_t^{'} \sim LN(R_t \sum_{\tau=0}^{t-1} i_\tau g_{t-\tau}, d),$$

$$i_t = (N - \sum_{j=0}^{t-1} i_j) \left(1 - \exp\left(-\frac{i_t^{'}}{N}\right)\right), \tag{2}$$

where the first equation makes infections latent and the second is an adjustment for population effects. In the above, N is the population size of that area, $LN(i_t^{'}, d)$ is a lognormal distribution with mean i_t and a standard deviation of $d\sqrt{i_t^{'}}$, with d being assigned a prior. The number of infections that would have occured if everyone in the population was susceptible is $i_t^{'}$.

The formula for i_t is derived from a continuous time model on $[t-1, t]$. This is to avoid discrete time effects such as infections going above the total population N. Specifically, we assume that the infections $i(s)$ in $[t-1, t-1+s]$ are given by the differential equation $\partial i(s)/\partial s = i_t'(1 - (\sum_{j=0}^{t-1} i_j + i(s))/N)$, which has the solution $i(1) = i_t$ as above. When we report reproduction numbers we adjust these for the population effect and report $R_t(1 - (\sum_{j=0}^{t-1} i_j)/N)$.

The model contains different observation types (cases, deaths, survey data), each of which will only be present at some time points (e.g. once per week). For an observation of type l at time t, we model the expected observation o_t^l as a weighted sum of past infections:

$$o_t^l = \alpha(t) \sum_{\tau=0}^{t-1} i_\tau \pi_{t-\tau}. \tag{3}$$

The observed data O_t^l are then noisy versions of this, i.e. $O_t^l \sim \mathcal{F}^l\left(o_t^l, \phi\right)$, where $\mathcal{F}$ is a distribution (e.g. Poission, negative binomial) parameterised by its mean o_t^l and possibly an auxiliary parameter ϕ.

Typical examples for observations include case or death counts. $\alpha(t)$ represents an ascertainment rate and $\pi_{t-\tau}$ represents the distribution that provides the weighting for past infections. For case or death data, $\alpha(t)$ would be the infection ascertainment rate (IAR) or the infection fatality rate (IFR) respectively, and $\pi_{t-\tau}$ would be the infection to case or the infection to death distribution respectively.

In our specific case, we parameterise $\mathcal{F}$ as negative-binomial distribution for modelling weekly deaths and cases data. To model seroprevalence data we parametrise $\mathcal{F}$ as a normal distribution. The exact details on how each different observation is modelled is given in supplementary information Section S1.

Interventions are not explicitly included in local models. Instead, we parameterise R_t with a random effect for each week of the epidemic, and for each LA separately (no
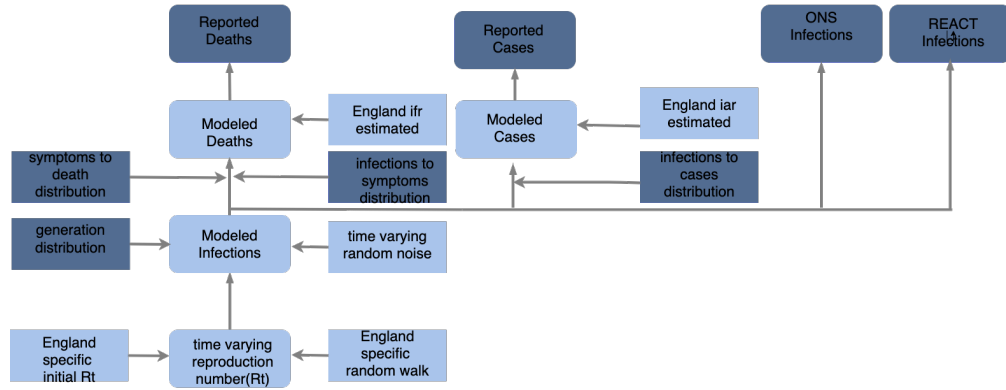
Fig. 1. Model diagram for first stage when we have all four different observations(deaths, cases, ONS infections, REACT infections) available for England and we use this to estimate time varying IFR and IAR. Dark blue nodes are observed.
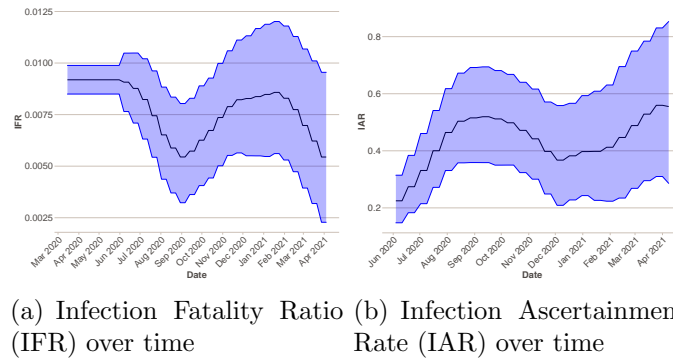

(a) Infection Fatality Ratio (IFR) over time

(b) Infection Ascertainment Rate (IAR) over time

Fig. 2. Time varying estimates of the infection fatality ratio (IFR), figure (a), and the infection ascertainment rate (IAR), figure (b), for England. The solid line is the mean estimate and the filled area denotes 90% pointwise credible intervals.

joint inference of parameters across different local authorities). Weekly random effects are encoded as a random walk with normally distributed updates. Thus, under the prior, each successive step the random effect has an equal chance of moving upwards or downwards from its current value. Our model is implemented using epidemia‡ (Scott et al., 2020), a general purpose R package for semi-mechanistic Bayesian modelling of epidemics built on top of RStan (Carpenter et al., 2017) and using an interface based on rstanarm (Goodrich et al., 2020).

We take a three-stage, top-down approach to fitting LA models.

We first calibrate the infection fatality ratio (IFR) and infection ascertainment rate (IAR) by fitting the model at national level (i.e. not broken down by region) to weekly deaths and cases in England, as well as to survey data from ONS and REACT. Figure
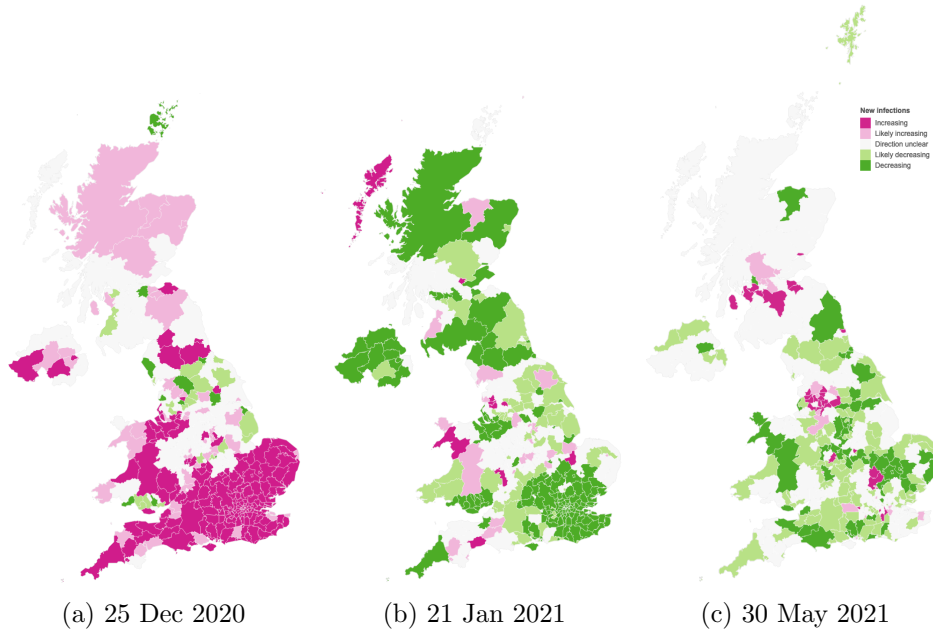
‡https://imperialcollegelondon.github.io/epidemia

(a) 25 Dec 2020   (b) 21 Jan 2021   (c) 30 May 2021

Fig. 3. Probability of epidemic growth by local area. We consider infections within a local area to be increasing if $R_t > 1$ with probability $\geq$ 90% (dark green), and likely increasing if $R_t > 1$ with probability between 75% and 90% (light green). Decreasing (dark pink) and likely decreasing (light pink) are defined symmetrically with $R_t < 1$.

1 gives an overview of this stage. For all subsequent individual models, we use these estimated IFR and IAR values (Figure 2) as priors.

In the second stage, we fit individual models to Northern Ireland, Scotland and Wales and to the 9 regions of England using death and case data aggregated to regional level from local authority level, using the priors estimated for the IFR and IAR in the first stage. The estimates from these 12 regions provide underlying trends of R_t for local areas within a particular region. We report the results of these regions, together with estimates for England and the UK as a whole. A diagram for this stage is in Figure S1 of the supplementary material.

Third, we fit individual models for each local area. Local values of R_t are parameterised as a sum of a weekly random walk and the estimated value of R_t from the region each local area is located in (multiplied by a tight prior around 1). There are three broad components to the model likelihood, arising from cases, deaths and seroprevalence. All model code is available at `https://github.com/ImperialCollegeLondon/covid19uklocal`, and detailed description of the model and fitting procedure are given in first the supplementary information Section S1 and in Bhatt et al. (2020); Scott et al. (2020). A diagram for this stage is in Figure S2 of the supplementary material.
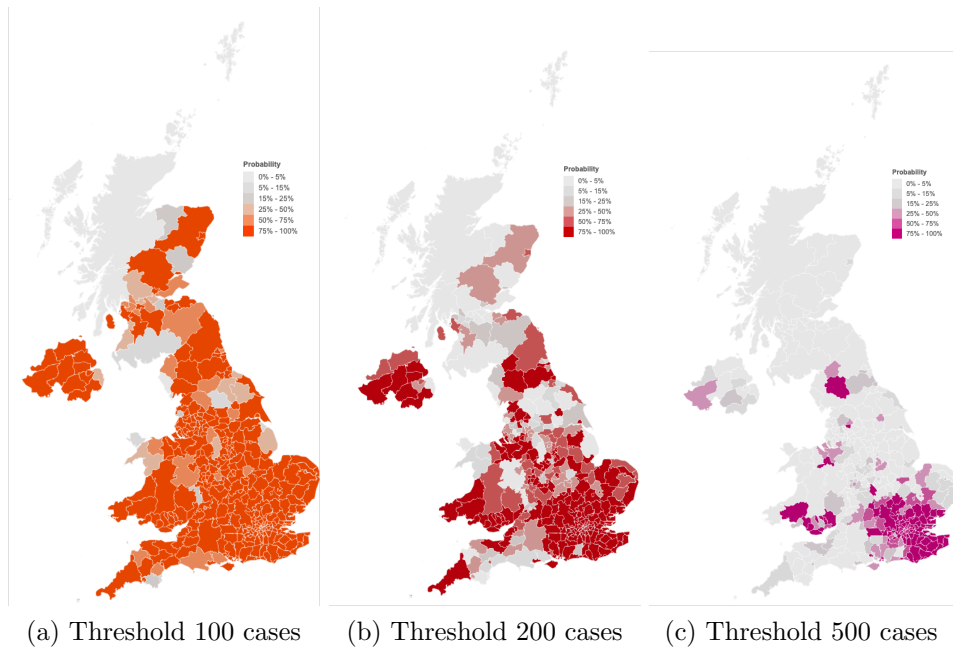
(a) Threshold 100 cases   (b) Threshold 200 cases   (c) Threshold 500 cases

Fig. 4. P(hotspot x) for different thresholds. Probability of local authority areas exceeding 100, 200 or 500 cases per 100K population for the period 27th Dec 2020 to 2nd Jan 2021 as projected on 25th Dec 2020.

4. Results

In this section we give some example outputs from the model.

Figure 3 shows the posterior probability that R_t is greater than 1 by local area for three different periods of epidemics, 25th Dec 2020, immediately before England was put into a national lockdown, 21st Jan 2021 just after the third lockdown, and 30th May 2021 three weeks before the planned removal of all restrictions. We include estimates of R_t and infections over time for each local area on our website§ (Gandy and Mishra, 2020), which is updated daily.

On our website, we have used the term "hotspot x" (e.g. "hotspot 50", "hotspot 100") which we define as a local authority, whose weekly reported cases per 100k population exceed x. We specifically project the probability of an area being a "hotspot x" in the next 3 weeks and report maps of these probabilities. Examples of such maps are shown in Figure 4 for the period 27th Dec 2020 to 2nd Jan 2021 as projected on 25th Dec 2020. As on our website, we have chosen different colours for different thresholds to set the plot for different thresholds clearly apart.

A monitoring system could use these probability projections to identify areas of concerns. They offer a combined view of both the current infections as well as the current growth of the epidemic (R_t) in a single number. A simplistic view that looks at either value in isolation can be misleading: for example, a low case count with high R_t,

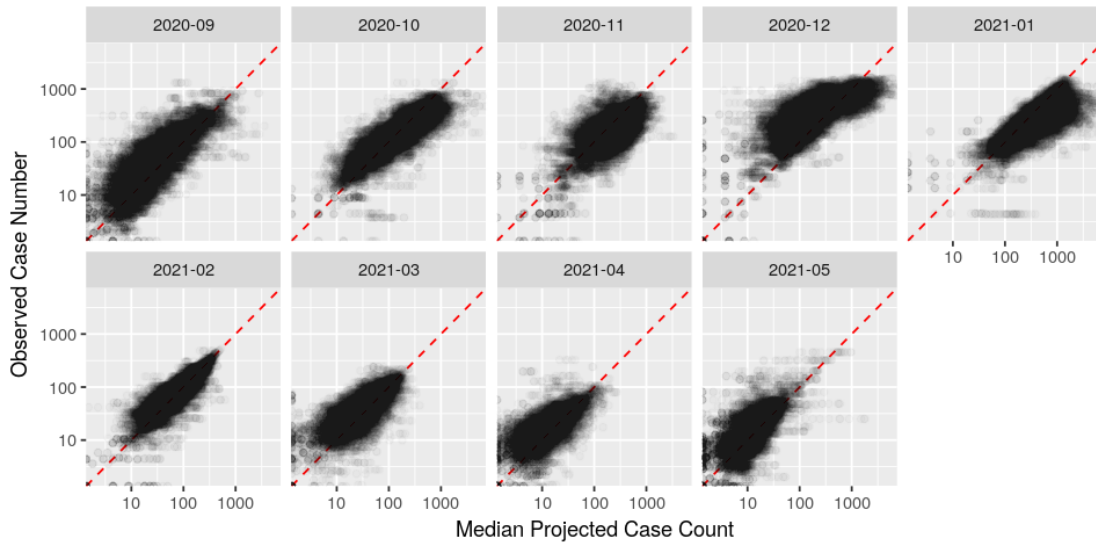§https://imperialcollegelondon.github.io/covid19local/

Fig. 5. Median projections versus observed weekly case numbers per 100k by month in which the projection was made

through exponential growth, can lead to a sudden increase in infections which can be more concerning than a high case count with (very) low R_t.

The tier monitoring system in Scotland used output from our model from 22 October 2020 onwards, specifically the hotspot probabilities (Scottish Government, 2020). Initially, our model was used on its own, from 25 Feb 2021 the outputs of this model were combined with the output from two other models.

5.   Model Evaluation

The model has been run regularly, mostly daily, since September 2020 and results were made publicly available. Projections were made for the upcoming three weeks. Figure 5 shows a comparison of the projections made by the model and the reported weekly case numbers per 100k population. The comparison is by month in which the projection was made. This is an out-of-sample test, as all of these projections were made before the case numbers were available.

Figure 5 shows a reasonable correspondence between projections and reported case numbers, in the sense that they seem to be correctly centered. Exceptions are partiuclarly in December 2020 and in January 2021. In December 2020, projections were generally lower than reality - this is most likely due to the emergence of the Kent variant of the virus, which has a substantially increased transmissibility (Volz et al., 2021). In January 2021 the projections were higher than the reported case counts - due to interventions being put into place nationwide. This is expected behaviour from the model - it projects the current state of the epidemic forward - it does not attempt to estimate the effect of policy changes or the emergence of a new variant of the virus.

6. Discussion

Our modelling approach has some underlying limitations. We have assumed homogeneous mixing of the population within local authorities and various age groups. All of our probability distributions for the delay between infection and symptom onset, between symptom onset and death, time between a person's infection and their subsequent transmission are assumed to be constant throughout.

Projections from the model assume no change in governmental interventions and human behaviour. Furthermore, interventions are not explicitly included in our models. Hence, the effect of a measure may not appear for 1-2 weeks after its implementation, once the random walk in R_t starts picking up a signal in the data. There are several reasons for this approach. First, reliable data on interventions and mobility is typically not immediately available. Even if it were available, the effects on transmission may only be identifiable over time. The random walk in R_t can pick up these effects automatically. Second, including interventions in the model would require frequent adjustments to the model, and would make providing daily updates difficult. A side effect of not including interventions is that potentially sharp transitions in R_t due to measures may be smoothed over time by the random walk.

An alternative formulation to the model would have the reproduction number R in (1) and/or the ascertainment rates α in (3) depending on the time of infection τ instead of the current time t. We chose the dependence on the current time t, as this allows the model to adapt to changes such as increased testing or non-pharmaceutical interventions that affect all infected individuals.

Apart from the estimates used from previous steps (e.g. IFR, IAR), we deliberately did not construct a joint model of all regions (e.g. using partial pooling in the spirit of Gelman (2006)). One reason is the computational demand for running such a model. A second reason is that we wanted the projections for one area to be not strongly affected by neighbouring areas, ensuring that decisions for an areas can be justified mostly by information from the same area.

The three stage modeling approach helps to robustly estimate the epidemic. The estimates of IFR and IAR from the first stage are reliable for England as they are inferred using serosurveys from both the ONS and REACT. In all remaining models, the individual IFR and IAR are specified by tight priors with the mean equal to the IFR and IAR for the whole of England. This step permits variation between individual models while also calibrating IFR/IAR against serosurveys. Recall that in the second stage we fit models for all regions in England, as well as for all nations in the UK. The LA models are then fit using the regional R_t (region for LTLAs in England and nations for local authorities in other nations) as a covariate for the local R_t in addition to the weekly random walk. The regional R_t trend helps to stabilise inference for local areas. Background regional R_t values are not used for the last 45 days to ensure that recent trends in local R_t are driven primarily by the data from the local area.

The renewal equation (1) propagates infections deterministically. This is generally suitable as infections become large, but in low incidence settings, estimation of R_t can be sensitive to random fluctuations and noise. This is why we treat infections as latent parameters which must be sampled. Infections are assigned a distribution with mean given by (1) and coefficient of variation d, which is assigned a prior. This extension

reflects a belief that changes in the number of infections at low infection counts provide limited evidence to ascertain R_t, and must be treated with caution. Supplementary information Eq (2) gives exact details of the formulation.

An important aspect of our framework is the projection of hotspots for all local areas. We argue that for understanding the true state of an epidemic in an area, only specifying R_t or current incidence is not enough. Summarising areas with R_t can be misleading as areas with low incidence might still be controlled with moderate measures. Similarly, high incidence areas that have $R_t < 1$ show signs of control and are not in an exponential growth phase. For these reasons, we believe that projections of future cases provide a more relevant indicator. Case projections combine the rate of change and the absolute values of infections into a single actionable number. Moreover, it is key to measure not just projected case counts but also the confidence the model has that those projections will be above a given threshold. Hence, in our framework we define hotspots based on the probability of a projection being greater than a specific threshold.

References

Scottish Health and Social Care Open Data (2020) Daily COVID-19 Cases in Scotland - Total Cases By Council Area -. URL: `https://www.opendata.nhs.scot/dataset/covid-19-in-scotland/resource/e8454cf0-1152-4bcb-b9da-4343f625dfef`.

Bhatt, S., Ferguson, N., Flaxman, S., Gandy, A., Mishra, S. and Scott, J. A. (2020) Semi-Mechanistic Bayesian Modeling of COVID-19 with Renewal Processes. *arXiv*.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. and Riddell, A. (2017) *Stan*: A Probabilistic Programming Language. *Journal of Statistical Software*, **76**, 1–32.

Department of Health and Social Care (2020) The R value and growth rate. URL: `https://www.gov.uk/guidance/the-r-value-and-growth-rate`. [Online; accessed 31. May 2021].

Department of Health, Northern Ireland (2020) COVID-19 - Daily Dashboard Updates. URL: `https://www.health-ni.gov.uk/articles/covid-19-daily-dashboard-updates`.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Imperial College COVID-19 Response Team, Ghani, A. C., Donnelly, C. A., Riley, S. M., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C. and Bhatt, S. (2020) Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*.

Gandy, A. and Mishra, S. (2020) ImperialCollegeLondon/covid19local: Website Release for Friday 20 Nov 2020. *Zenodo*.

Gelman, A. (2006) Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. *Technometrics*, **48**, 432–435.

Goodrich, B., Gabry, J., Ali, I. and Brilleman, S. (2020) rstanarm: Bayesian applied regression modeling via Stan. URL: `https://mc-stan.org/rstanarm`. R package version 2.21.1.

Liu, Y., Gayle, A., Wilder-Smith, A. and Rocklöv, J. (2020) The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*.

Mishra, S., Berah, T., Mellan, T. A., Unwin, H. J. T., Vollmer, M. A., Parag, K. V., Gandy, A., Flaxman, S. and Bhatt, S. (2020) On the derivation of the renewal equation from an age-dependent branching process: an epidemic modelling perspective.

Office for National Statistics (2020a) Coronavirus (COVID-19) Infection Survey - Office for National Statistics. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata`.

— (2020b) Death registrations and occurrences by local authority and health board. URL: `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/datasets/deathregistrationsandoccurrencesbylocalauthorityandhealthboard`.

Public Health England (2020) Official UK Coronavirus Dashboard. URL: `https://coronavirus.data.gov.uk/details/cases`.

Public Health Wales (2020) Rapid COVID-19 virology - Public. URL: `https://public.tableau.com/profile/public.health.wales.health.protection#!/vizhome/RapidCOVID-19virology-Public/Headlinesummary`.

REACT Study (2020) Real-time Assessment of Community Transmission findings. URL: `https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/real-time-assessment-of-community-transmission-findings`.

Riley, S., Ainslie, K. E. C., Eales, O., Jeffrey, B., Walters, C. E., Atchison, C. J., Diggle, P. J., Ashby, D., Donnelly, C. A., Cooke, G., Barclay, W., Ward, H., Taylor, G., Darzi, A. and Elliott, P. (2020) Community prevalence of SARS-CoV-2 virus in England during may 2020: REACT study. *medRxiv*.

Scott, J. A., Gandy, A., Mishra, S., Unwin, J., Flaxman, S. and Bhatt, S. (2020) epidemia: Modeling of epidemics using hierarchical Bayesian models. URL: `https://imperialcollegelondon.github.io/epidemia/`. R package version 0.5.3.

Scottish Government (2020) Coronavirus (COVID-19): modelling the epidemic - gov.scot. URL: `https://www.gov.scot/collections/coronavirus-covid-19-modelling-the-epidemic`.

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G. T., Fu, H., Dighe, A., Jamie, T., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Laydon, D. and Nedjati, G. (2020) Estimates of the severity of COVID-19 disease. *Lancet Infect Dis*.

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., Hinsley, W. R., Laydon, D. J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C. V., Boyd, O., Loman, N. J., McCrone, J. T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D. K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D. P., COVID-19 Genomics UK (COG-UK) consortium, Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A. and Ferguson, N. M. (2021) Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*.

Walker, P. G., Whittaker, C., Watson, O. J., Baguelin, M., Winskill, P., Hamlet, A., Djafaara, B. A., Cucunubá, Z., Mesa, D. O., Green, W. et al. (2020) The impact of COVID-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*, **369**, 413–422.

Supplementary Material

S1.   Model details

We provide a compact description of the the model. More details are available in Bhatt et al. (2020).

Both deaths and cases are observed in our model. We define weekly deaths, D_w, for weeks $w \in \{1, \ldots, w(n)\}$, where n is the total number of days in the data. These weekly deaths are modelled using a positive real-valued function $d_w = \mathbb{E}[D_w]$ that represents the expected number of deaths attributed to COVID-19. The weekly deaths D_w are assumed to follow a negative binomial distribution with mean d_w and variance $d_w + \frac{d_w^2}{\psi_1}$, where ψ_1 follows a positive half normal distribution, i.e.

$$D_w \sim \text{Negative Binomial}\left(d_w, d_w + \frac{d_w^2}{\psi_1}\right), \quad t = 1, \ldots, n \tag{S1}$$

$$\psi_1 \sim \mathcal{N}^+(0, 40). \tag{S2}$$

Here, $\mathcal{N}(\mu, \sigma)$ denotes a normal distribution with mean μ and standard deviation σ. We say that X follows a positive half normal distribution $\mathcal{N}^+(0, \sigma)$ if $X \sim |Y|$, where $Y \sim \mathcal{N}(0, \sigma)$.

We link our observed deaths mechanistically to transmission as in Flaxman et al. (2020). We use a previously estimated COVID-19 specific distribution of times from infection to death π, as detailed in Verity et al. (2020); Walker et al. (2020). We assume the distribution of times from infection to death π (infection-to-death) to be the convolution of an infection-to-onset distribution (π') (Walker et al., 2020) and an onset-to-death distribution (Verity et al., 2020):

$$\pi \sim \text{Gamma}(5.1, 0.86) + \text{Gamma}(17.8, 0.45). \tag{S3}$$

Similar to generation distribution, π is further discretized via $\pi_s = \int_{s-0.5}^{s+0.5} \pi(\tau)d\tau$ for $s = 2, 3, \ldots$, and $\pi_1 = \int_0^{1.5} \pi(\tau)d\tau$, where $\pi(\tau)$ is the density of π.

For estimating time varying infection fatality ratio (IFR, probability of death given infection) we use the estimates from Riley et al. (2020) as a starting point and then a bi-weekly change is estimated using the infections survey data from Office for National Statistics (2020a) and REACT Study (2020).

The expected number of deaths d_w, in a given week w, is given by the following discrete sum:

$$d_w = \sum_{\tau=0}^{t-1} \text{IFR}(w)i_\tau \pi_{t-\tau}^W, \tag{S4}$$

where t is the total number of days till week w, i_τ is the number of new infections on day τ and where π^W is the transformation of discretized π to a weekly lag.

We also observe weekly cases C_w from 1st June 2020 $w \in \{w(t = 2020\text{-}06\text{-}01), \ldots, w(n)\}$. Similar to weekly deaths, weekly cases are modelled using a positive real-valued function $c_w = \mathbb{E}[C_w]$ that represents the expected number of cases identified in testing for COVID-19. Again, the weekly cases C_w are assumed to follow a negative binomial distribution but with mean c_w and variance $c_w + \frac{c_w^2}{\psi_2}$, where ψ_2 follows a positive half normal

distribution, i.e.

$$C_w \sim \text{Negative Binomial}\left(c_w, c_w + \frac{c_t^2}{\psi_2}\right), \quad t = t_c, \ldots, n, \tag{S5}$$

$$\psi_2 \sim \mathcal{N}^+(0, 40). \tag{S6}$$

We assume the lag between an infection to be identified as a case, π^{i2c}, is zero for first three days and then has an equal chance over next 10 days, i.e.

$$\pi^{i2c} = c\left(0, 0, 0, repeat(\frac{1}{10}, 10)\right). \tag{S7}$$

We link the observed weekly cases and estimated daily infections using an estimated parameter infection ascertainment ratio (IAR), which is defined as the number of reported cases divided by the true number of infections (including both symtomatic and asymptomatic infections).

The expected number of cases c_w, in a given week w is given by the following discrete sum:

$$c_w = \sum_{\tau=t-13}^{t-1} \text{IAR}(w) i_\tau \pi_{t-\tau}^{i2c}, \tag{S8}$$

where, again, where t is the total number of days till week w, and i_τ is the number of new infections on day τ.

We also observe daily infections from ONS I_t^{ons} in our model. ONS infections are modelled using a positive real-valued function $i_t^{ons} = \mathbb{E}[I_t^{ons}]$ that represents the expected number of infections given by the ONS infection survey. Daily infections from ONS I_t^{ons} are assumed to follow a normal distribution but with mean i_t^{ons} and standard deviation given by the 95% CI in ONS infection survey, i.e.

$$I_t^{ons} \sim \mathcal{N}\left(i_t^{ons}, \frac{95\%CI}{5}\right), \quad t = t_c, \ldots, n, \tag{S9}$$

$$\tag{S10}$$

The expected number of ONS infections i_t^{ons}, on a given day t is given :

$$i_t^{ons} = i_t, \tag{S11}$$

We also observe total infections from REACT $I_{t,total}^{react}$ in our model. REACT infections are modelled using a positive real-valued function $i_{t,total}^{react} = \mathbb{E}[I_{t,total}^{react}]$ that represents the expected total number of infections given by the REACT. Total infections from REACT $I_{t,total}^{react}$ are assumed to follow a normal distribution but with mean $i_{t,total}^{react}$ and standard deviation given by the 95% CI in REACT survey, i.e.

$$I_{t,total}^{react} \sim \mathcal{N}\left(i_{t,total}^{react}, \frac{95\%CI}{5}\right), \quad t = t_c, \ldots, n, \tag{S12}$$

$$\tag{S13}$$

The expected number of total infections $i_{t,total}^{react}$, on a given day t is given :

$$i_{t,total}^{react} = \sum_{\tau=0}^{t} i_\tau, \tag{S14}$$

We parametrise $R_{t,m}$ with a random effect for each week of the epidemic as follows

$$R_t = R_0 \cdot f\left(-\epsilon_{w(t)}\right), \tag{S15}$$

where $f(x) = 2\exp(x)/(1+\exp(x))$ is twice the inverse logit function and $\epsilon_{w(t)}$ is a weekly random walk (RW) process, that captures variation between R_t in each subsequent week.

Following Liu et al. (2020), the prior distribution for R_0 was chosen to be

$$R_0 \sim \mathcal{N}(3.28, 0.5) \tag{S16}$$

We assume that seeding of new infections begins 30 days before the day after a state has cumulatively observed 10 deaths. From this date, we seed our model with 6 sequential days of an equal number of infections: $i_1 = \cdots = i_6 \sim \text{Exponential}(\frac{1}{\tau})$, where $\tau \sim \text{Exponential}(0.03)$. These seed infections are inferred in our Bayesian posterior distribution.

The weekly effect is modelled as a weekly random walk process, centred around 0 with variance σ_ϵ that, starts on the first day of its seeding of infections, i.e. 30 days before a total of 10 cumulative deaths have been observed in this state. The RW process starts with $\epsilon_1 = 0$,

$$\vec{\epsilon}_{raw} \sim \mathcal{N}(0, 1) \tag{S17}$$

$$\epsilon_{w(t)} = \left(\sum_{i=1}^{|\vec{\epsilon}_{raw}|} \epsilon_{raw}[i]\right) * \sigma_\epsilon \text{ for } w(t) = 2, 3, 4, \ldots \tag{S18}$$

The prior for σ_ϵ, the variance of RW process ϵ, is chosen as $\sigma_\epsilon \sim \mathcal{N}^+(0, .2)$. The conversion from days to weeks is encoded in $w(t)$. Every 7 days, w is incremented, i.e. we set $w(t) = \lfloor(t - t^{start})/7\rfloor + 1$, where t^{start} is the first day of seeding.

We estimated parameters independently for each local authority. Fitting was performed with the R package epidemia (Scott et al., 2020), written in the probabilistic programming language Stan (Carpenter et al., 2017) using an adaptive Hamiltonian Monte Carlo (HMC) sampler.

Now for running our model not all observations are available at all levels. Hence we take a three stage approach to estimate the parameters as robustly as possible.

In first stage, Figure 1,we fit a model to entire of Engalnd, where we have all observations available, i.e., cases, deaths, ONS infections, REACT attack rate (total cumulative population infected). This means while fitting all four observations are used for inferring parameters namely via the observation process defined in Eq (S1), Eq (S5), Eq (S9), and Eq (S12). This step is very crucial as it is the only place we have an observation for infections. Hence, we estimate $ifr(w)$ and $iar(w)$ only in this step, in all other steps we use provide tight priors around the estimates from this step.
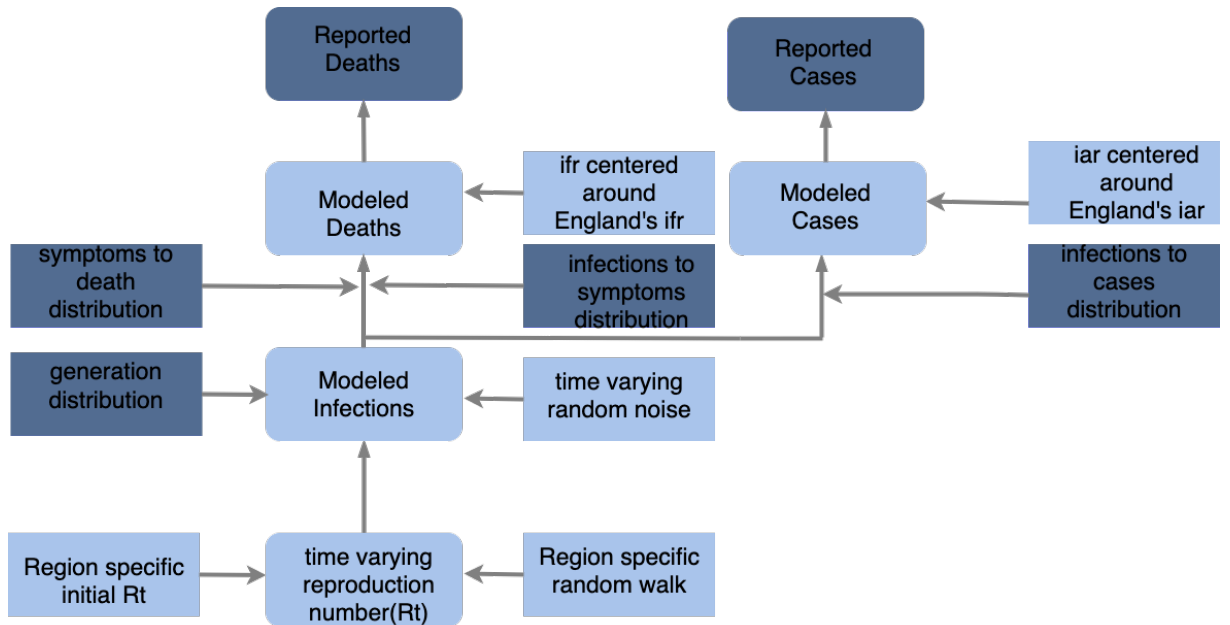
Fig. S1. Model diagram for second stage for regions of England, as well as all nations of the UK. Dark blue nodes are observed.

In second stage, Figure S1, we fit individual models to all regions in England, as well as to all nations in the UK. The observations used in this stage are deaths and cases, hence the likelihood of model in this stage is calculated by adding Eq (S1) and Eq (S5). As stated earlier the $ifr(w)$ and $iar(w)$ used in Eq (S1) and Eq (S5) are provided strong priors centred around the values estimated in first stage.

Finally, in our last stage, Figure S2, we fit individual models to all local authorities (LTLAs in England, local authorities in all other nations of the UK). As in second stage the likelihood of model is only composed of Eq (S1) and Eq (S5). Additionally in this stage we use the R_t of the region (region for LTLAs in England and nations for local authorities in other nations) as a covariate for R_t of the local area in addition to the weekly random walk.
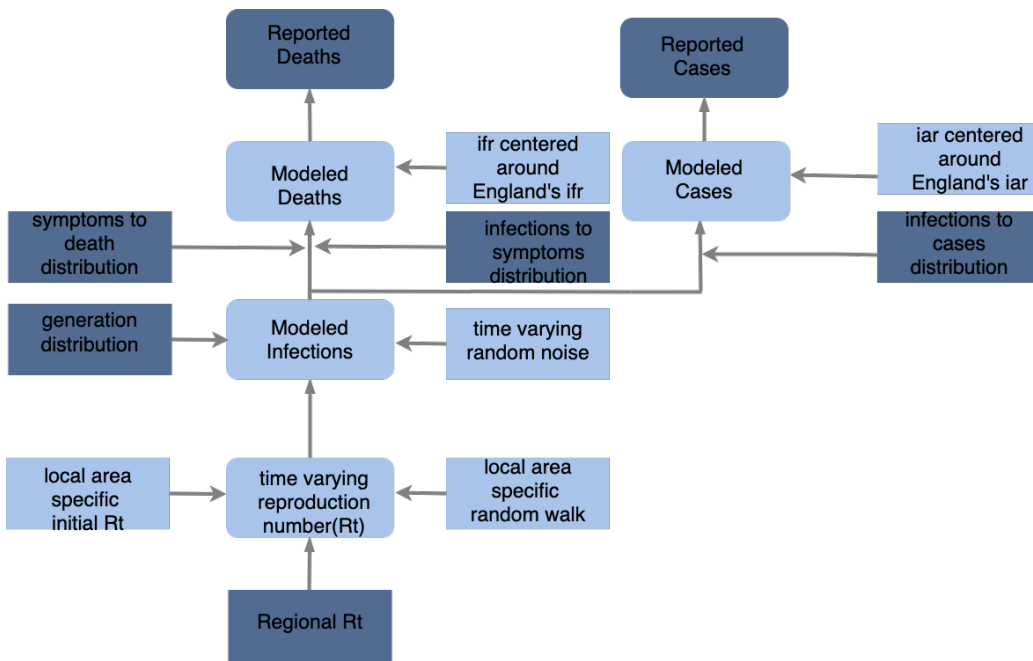
Fig. S2. Model diagram for local authorities epidemics modelling. Dark blue nodes are observed.