

COVID-19 Bayesian Spatial Stochastic Model-based Risk Estimation

Chris Jewell¹, Jonathan Read¹, Gareth Roberts², Barry
Rowlingson¹, Chris Suter³

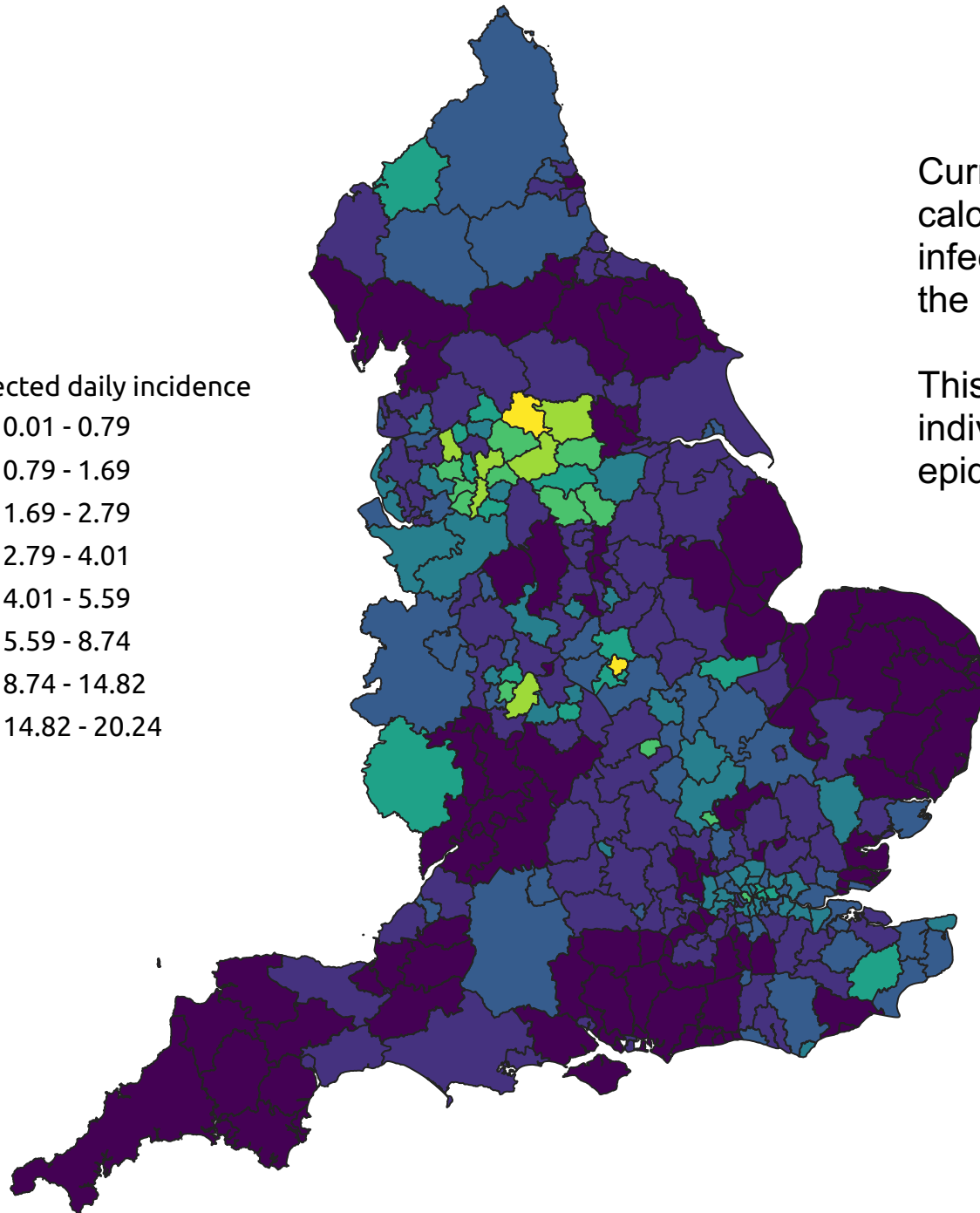
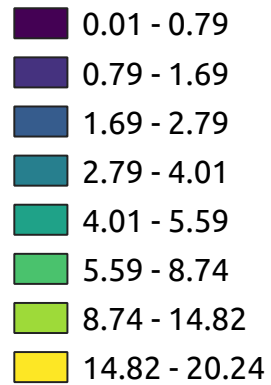
¹CHICAS, Lancaster University; ²Department of Statistics, University of Warwick; ³Google Brain, New York.

3rd August 2020

Spatial COVID-19 risk estimation for 30th July 2020, given PHE Anonymised Line Listing Data (Pillar 1 and 2) between 1st May 2020 and 30th July 2020.

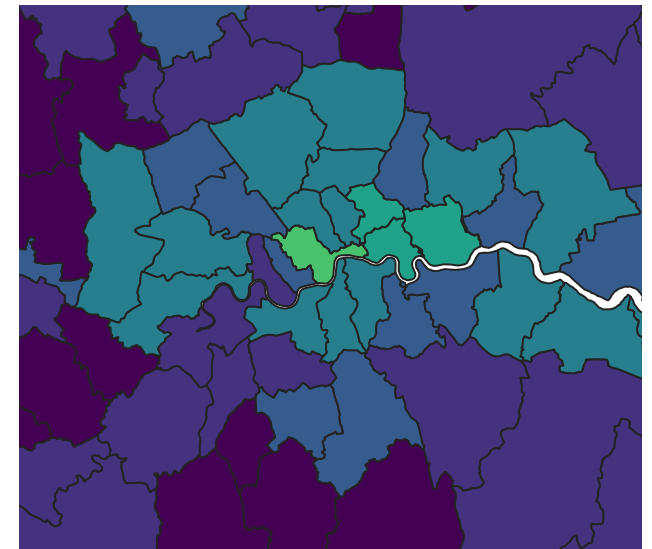
Current daily absolute incidence per LTLA

Expected daily incidence



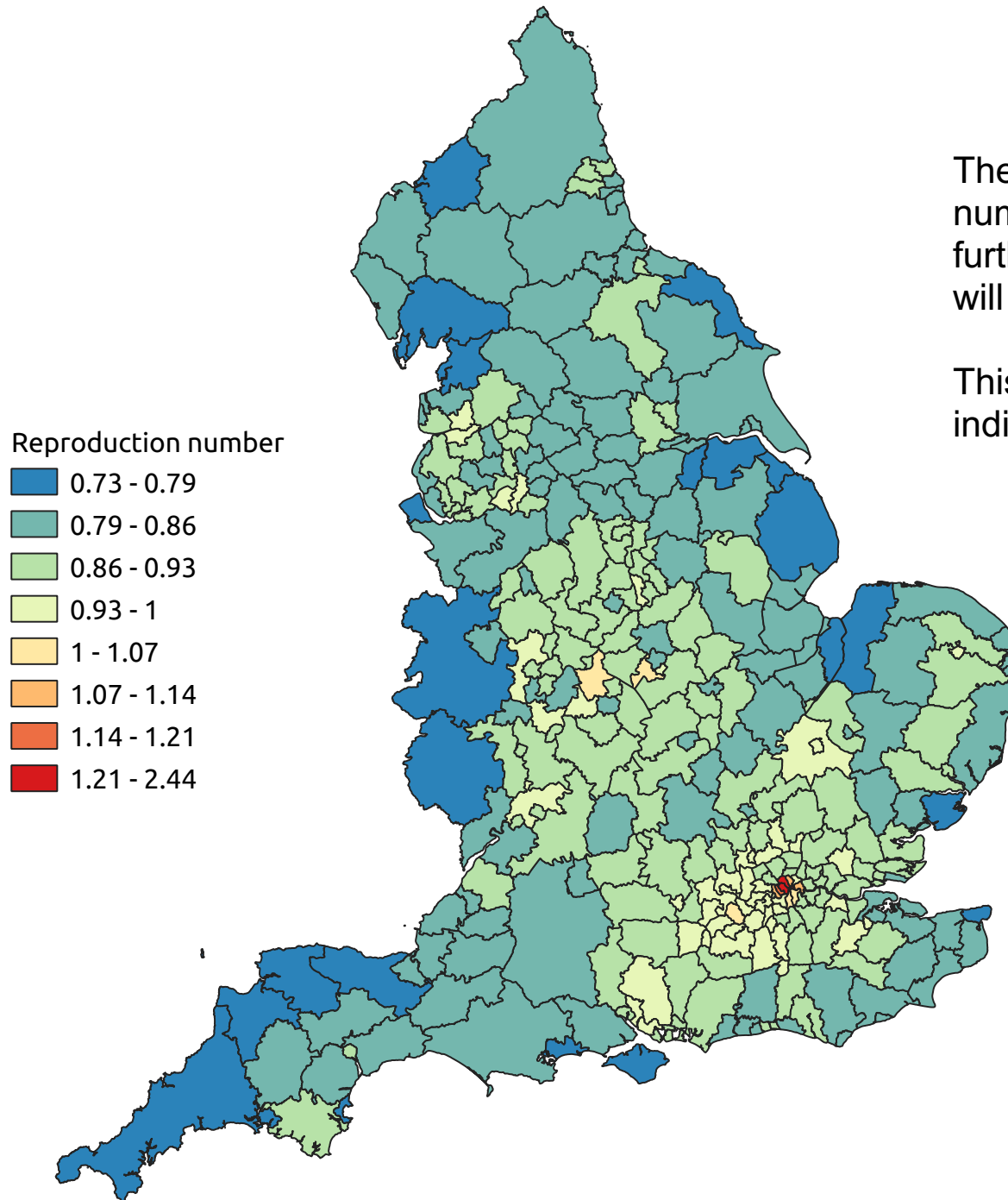
Current daily absolute incidence is calculated as the current daily force of infection on an individual multiplied by the population size for that LTLA.

This is a measure of current risk **to** individuals from the **current** state of the epidemic.



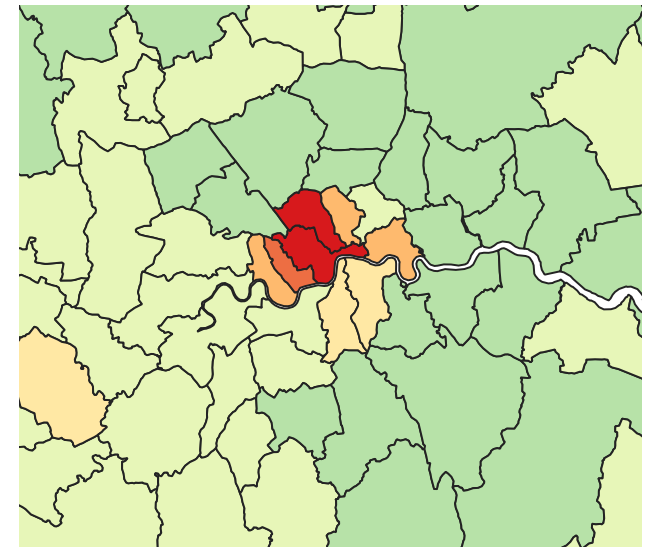
Inner London

Current LTLA-specific reproduction number



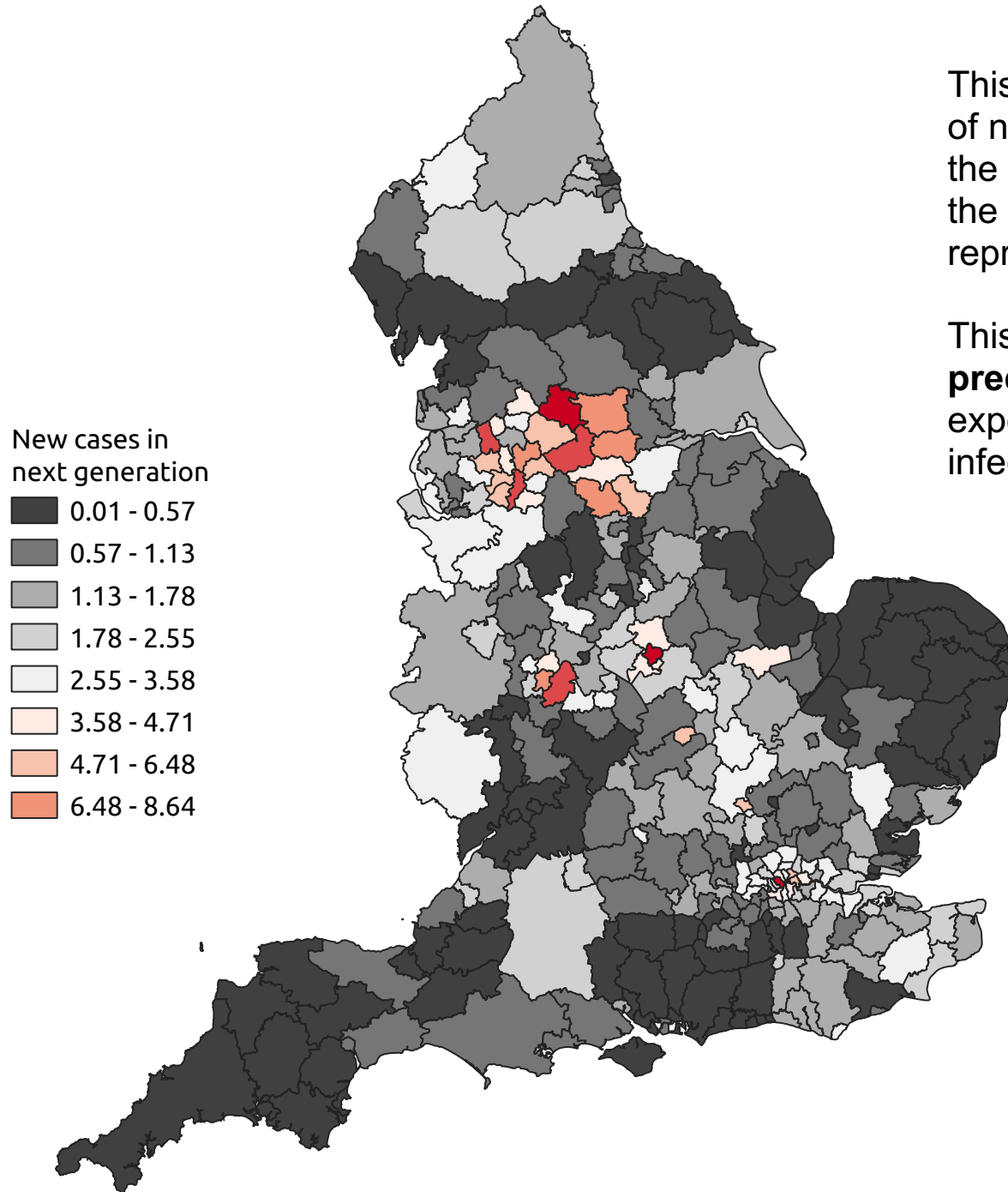
The current LTLA-specific reproduction number gives the expected number of further infections an infected individual will cause within each LTLA.

This is a measure of risk **from** individuals **if** they become infected.



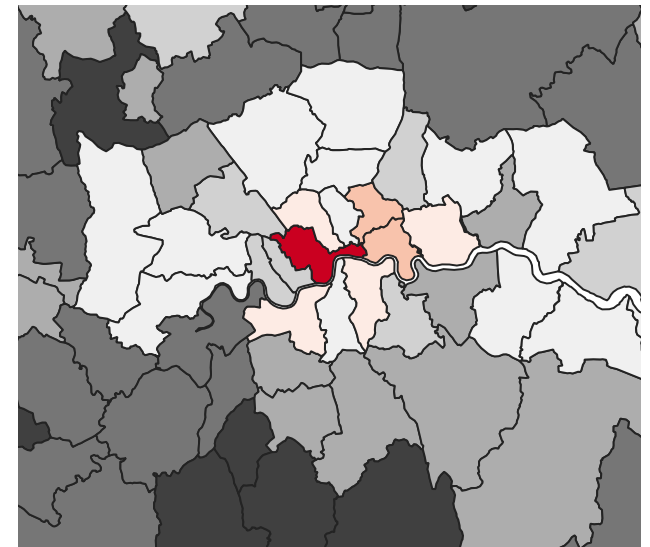
Inner London

LTLA specific expected new cases



This maps shows the expected number of new cases **in the next generation** of the disease. It is obtained by multiplying the current incidence by the reproduction number.

This is a measure of the overall **predicted** case load in each LTLA we expect within the next generation of infection (approximately 1 week).



Inner London

Description of epidemic modelling approach

State transition model

A discrete time stochastic SEIR model, with a time step of 1 day to match the frequency at which case data is provided by PHE. We assume the state evolves according to a discrete-time Markov process (i.e. chain-binomial sampling).

Population mixing model

We assume a "meta-population" structure, where individuals within each of 315 Local Authority Districts in England mix and transmit infection homogeneously. Transmission between regions is assumed to follow a symmetric mixing matrix determined from commuting frequency observed by Census 2011 as well as time-varying traffic volume data from DfT. For each meta-population, we take population estimates from ONS December 2019 population predictions.

Event data

We assume that in our model I->R transition events represent individuals testing positive for COVID-19 through Pillar 1 and Pillar 2 tests (and thereafter self-isolating with 100% efficacy). We assume that both the S->E and E->I transitions are censored.

Parameters

We assume three unknown parameters: a) a fortnightly-varying baseline transmission rate implemented as a Gaussian process with Matern correlation; b) the relative contribution of inter-LAD transmission versus within-LAD transmission; the mean infectious period (i.e. time spent in state I). We assume a known mean latent period of 2 days.

Parameter inference

The model is trained on COVID-19 case data geolocated to LAD level between 1st May 2020 and 30th July 2020 inclusive. Fitting is performed using novel data-augmentation MCMC methodology to account for the censored S->E and E->I transition events (manuscript in preparation). The algorithm is implemented in Python 3.7 using Tensorflow 2.3 and Tensorflow Probability 0.11 machine learning libraries. Markov chains are inspected visually for ergodic convergence, with suitable burn-in iterations removed and chain-thinning performed.

Mapable metrics

Current case incidence is calculated as the posterior expected force of infection on an individual within a LAD on 30th July 2020 (including "occult" [unobserved] exposed and infected individuals) multiplied by the LAD population size to give an estimate of new cases per day. LAD-specific reproduction number is calculated as the posterior expected number of further infections each infected individual will give rise to nationally (i.e. both within- and between-LAD infections). The LTLA-level expected number of new cases in the next generation is calculated as the produce of the incidence and reproduction number, and gives a simple metric for the national-level risk posed to the country by the current outbreak status in each LAD.