COVID-19 Bayesian Spatial Stochastic Model-based Risk Estimation

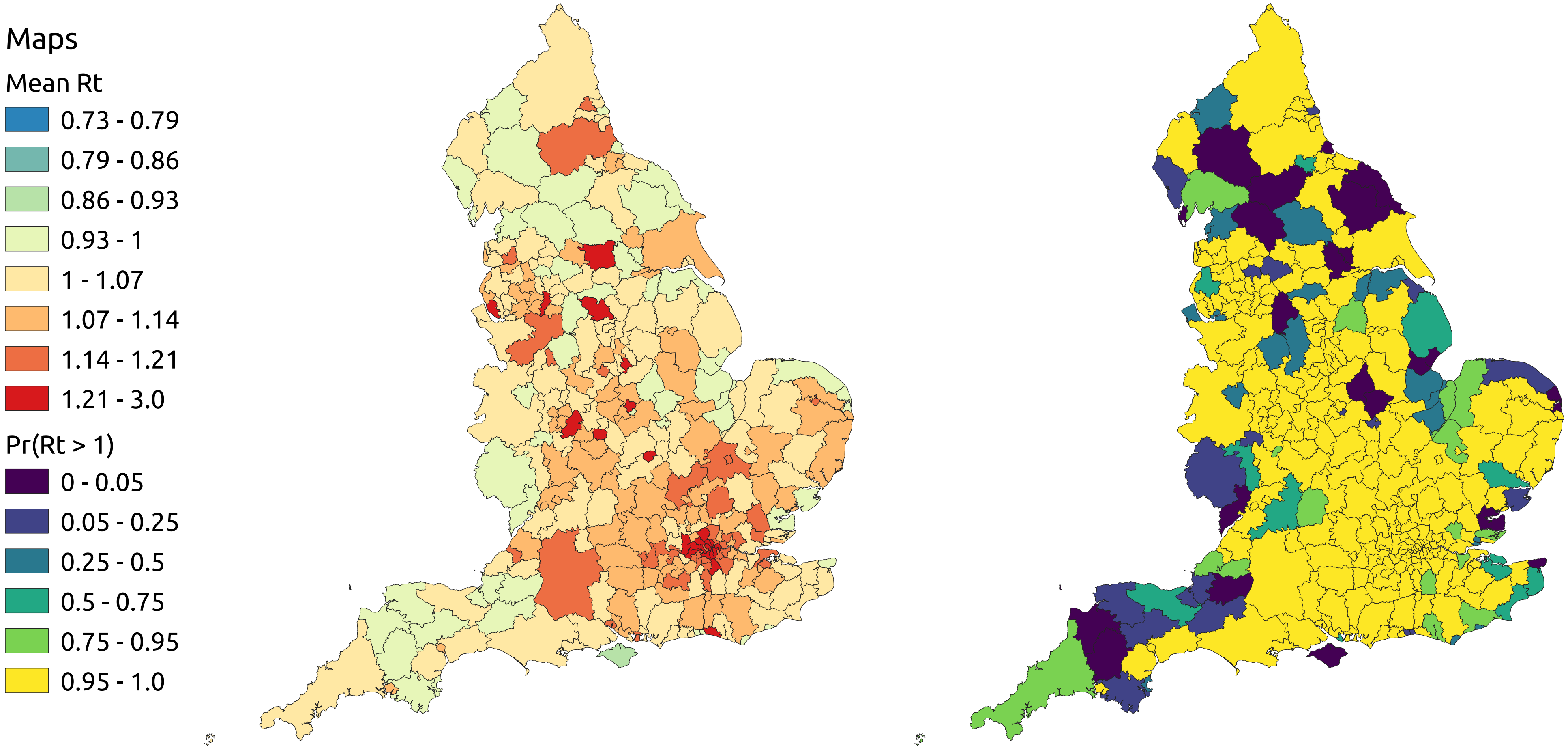
Chris Jewell[1], Jonathan Read[1], Gareth Roberts[2], Barry Rowlingson[1], Christopher Suter[3]

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

8th September 2020

Spatial COVID-19 risk estimation for 4th September 2020, given PHE Anonymised Line Listing Data (Pillar 1 and 2) between 12th June 2020 and 3rd September 2020 inclusive.

Spatial reproduction number at 2020-09-04

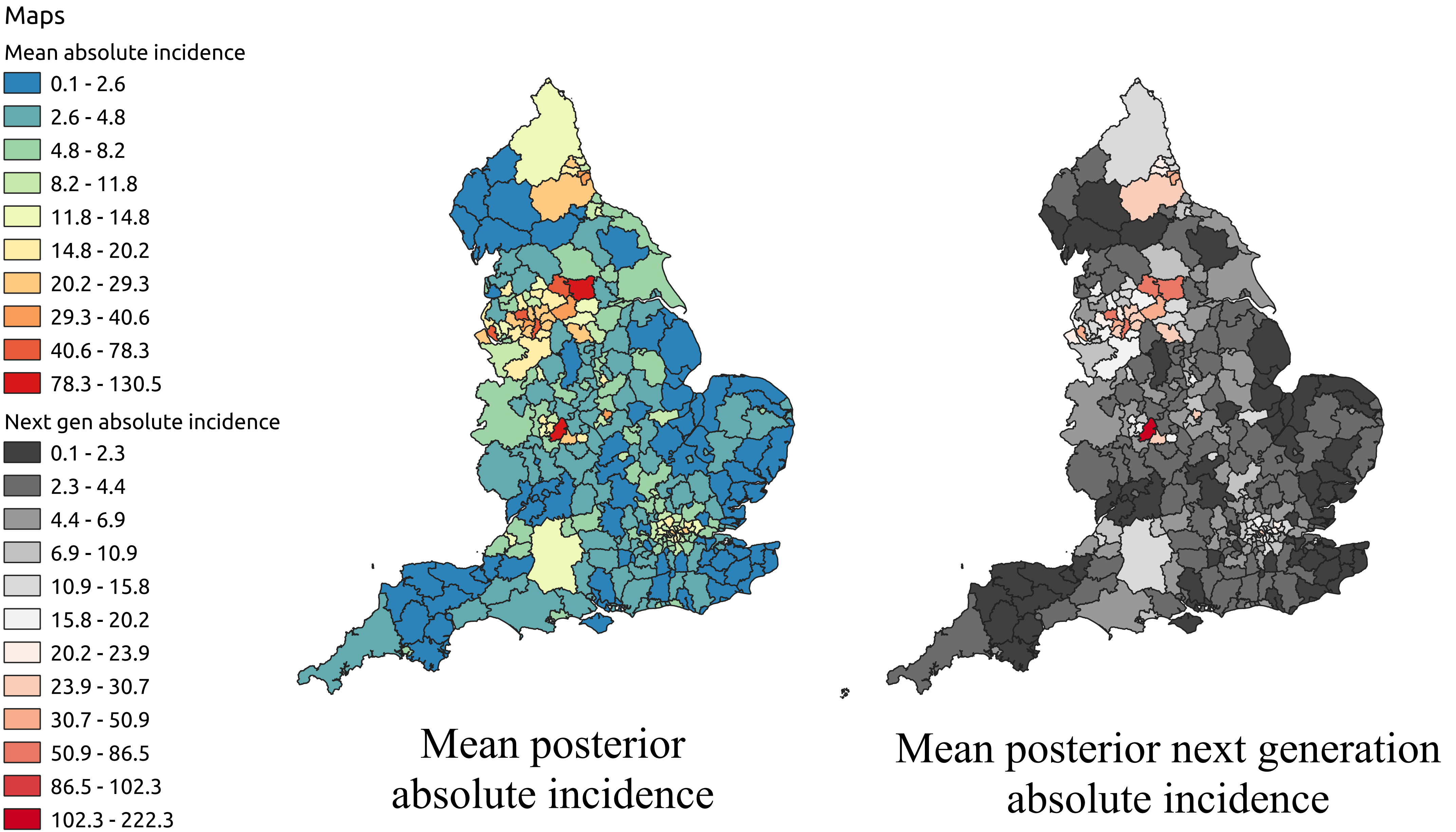


Mean posterior Rt

Pr(Rt > 1)

England-wide Rt = 1.29
95% credibility interval (1.26, 1.33)

Absolute incidences as of 2020-09-04



Top 20 LADs sorted by Posterior Mean Rt

LAD Code	LAD Name	Rt	Pr(Rt > 1)	Current cases per day	Next gen cases
E09000001,E09000033	City of London:Westminster	3.08 (2.96, 3.22)	1.00	28.04 (24.09, 32.77)	36.23 (30.8, 42.8)
E08000025	Birmingham	1.7 (1.66, 1.76)	1.00	130.46 (120.62, 142.23)	168.49 (153.11, 187.09)
E09000007	Camden	1.55 (1.51, 1.6)	1.00	12.5 (9.92, 15.07)	16.14 (12.87, 19.53)
E09000030	Tower Hamlets	1.49 (1.46, 1.54)	1.00	17.23 (13.86, 21.52)	22.26 (17.86, 28.01)
E08000003	Manchester	1.46 (1.43, 1.51)	1.00	69.9 (63.35, 78.45)	90.26 (81.57, 102.33)
E06000018	Nottingham	1.44 (1.4, 1.48)	1.00	18.12 (14.58, 23.15)	23.4 (18.74, 29.99)
E06000016	Leicester	1.44 (1.4, 1.48)	1.00	29.28 (22.82, 34.78)	37.82 (29.3, 44.89)
E09000028	Southwark	1.4 (1.36, 1.44)	1.00	12.49 (9.26, 16.12)	16.13 (11.91, 21.21)
E09000017	Hillingdon	1.36 (1.33, 1.4)	1.00	9.76 (6.88, 13.39)	12.61 (8.86, 17.56)
E09000019	Islington	1.34 (1.31, 1.39)	1.00	10.03 (7.22, 13.51)	12.95 (9.45, 17.25)
E09000022	Lambeth	1.33 (1.3, 1.37)	1.00	16.51 (12.88, 20.23)	21.33 (16.41, 26.68)
E09000032	Wandsworth	1.33 (1.3, 1.37)	1.00	12.91 (9.7, 16.26)	16.67 (12.56, 21.31)
E09000018	Hounslow	1.29 (1.26, 1.33)	1.00	11.13 (8.15, 14.49)	14.37 (10.48, 18.92)
E08000035	Leeds	1.28 (1.25, 1.32)	1.00	78.33 (70.61, 86.39)	101.15 (90.16, 112.46)
E09000009	Ealing	1.27 (1.24, 1.3)	1.00	13.95 (10.91, 16.68)	18.01 (14.2, 21.66)
E08000012	Liverpool	1.25 (1.23, 1.29)	1.00	40.63 (35.01, 46.47)	52.44 (45.24, 59.51)
E09000003	Barnet	1.25 (1.22, 1.29)	1.00	17.64 (13.95, 21.54)	22.77 (18.13, 27.63)
E09000013	Hammersmith and Fulham	1.24 (1.22, 1.28)	1.00	9.89 (7.24, 12.75)	12.77 (9.29, 16.55)
E09000008	Croydon	1.23 (1.21, 1.27)	1.00	11.15 (7.65, 14.87)	14.41 (9.74, 19.2)
E08000026	Coventry	1.23 (1.2, 1.26)	1.00	18.25 (14.01, 23.36)	23.56 (18.21, 30.32)

Remarks

- Values shown as x (y, z) represent posterior mean and 95% credibility intervals.
- "Next gen cases" gives the expected number of further cases in the next generation of infection, multiplying Rt by the current number of infected individuals.
- The top 20 LADs ranked by mean Rt are shown above. All such LADs correspond to highly populated and highly connected urban centres.
- Rt measures the potential for a LAD to create more cases both within itself and in the rest of the country. Thus LADs with a high Rt may be considered high risk *if* they develop a high number of cases.
- Note that is it currently LADs with *both* a high Rt and high mean absolute incidence that pose the greatest ongoing case risk.

Top 20 LADs sorted by expected number of cases in the next generation of infection

LAD Code	LAD Name	Rt	Pr(Rt > 1)	Cases per day	Next gen cases
E08000025	Birmingham	1.7 (1.66, 1.76)	1.00	130.46 (120.62, 142.23)	168.49 (153.11, 187.09)
E08000035	Leeds	1.28 (1.25, 1.32)	1.00	78.33 (70.61, 86.39)	101.15 (90.16, 112.46)
E08000003	Manchester	1.46 (1.43, 1.51)	1.00	69.9 (63.35, 78.45)	90.26 (81.57, 102.33)
E08000032	Bradford	1.12 (1.1, 1.16)	1.00	69.04 (61, 76.16)	89.14 (78.95, 99.42)
E08000001	Bolton	1.09 (1.06, 1.12)	1.00	67.91 (59.37, 77.97)	87.66 (76.7, 100.42)
E08000012	Liverpool	1.25 (1.23, 1.29)	1.00	40.63 (35.01, 46.47)	52.44 (45.24, 59.51)
E08000034	Kirklees	1.05 (1.03, 1.08)	1.00	37.91 (31.33, 43.25)	48.94 (40.72, 55.91)
E08000024	Sunderland	1.07 (1.05, 1.1)	1.00	35.61 (31.72, 39.62)	45.97 (40.83, 51.41)
E08000006	Salford	1.12 (1.09, 1.15)	1.00	35.5 (29.83, 42.45)	45.86 (38.27, 56.02)
E06000016	Leicester	1.44 (1.4, 1.48)	1.00	29.28 (22.82, 34.78)	37.82 (29.3, 44.89)
E08000019	Sheffield	1.22 (1.19, 1.26)	1.00	29.22 (23.24, 36.15)	37.72 (29.87, 46.6)
E08000005	Rochdale	1.05 (1.03, 1.08)	1.00	28.26 (23.11, 33.74)	36.5 (29.8, 43.79)
E09000001,E09000033	City of London:Westminster	3.08 (2.96, 3.22)	1.00	28.04 (24.09, 32.77)	36.23 (30.8, 42.8)
E08000004	Oldham	1.03 (1.01, 1.06)	1.00	27.95 (22.59, 34.53)	36.08 (28.88, 44.29)
E08000008	Tameside	1.04 (1.02, 1.07)	1.00	26.8 (21.47, 32.71)	34.6 (27.68, 41.8)
E06000047	County Durham	1.2 (1.17, 1.23)	1.00	25.63 (21.47, 31.29)	33.08 (27.7, 39.99)
E08000029	Solihull	1.1 (1.07, 1.13)	1.00	24.69 (20.09, 30.1)	31.87 (25.62, 38.75)
E08000009	Trafford	1.11 (1.09, 1.15)	1.00	24.33 (20.09, 28.43)	31.42 (26.01, 37.09)
E08000015	Wirral	1 (0.98, 1.03)	0.47	23.67 (18.86, 28.88)	30.56 (24.13, 37.39)
E08000021	Newcastle upon Tyne	1.19 (1.17, 1.23)	1.00	23.36 (19.23, 28.13)	30.17 (24.84, 36.92)

Remarks

- Values shown as x (y, z) represent posterior mean and 95% credibility intervals.
- "Next gen cases" gives the expected number of further cases in the next generation of infection, multiplying Rt by the current number of infected individuals.
- This table ranks LADs by the expected number of further cases in the next generation of infection, calculated as the product of Rt and the mean absolute incidence.
- The top 20 LADs are expected to represent approximately 40% of the total case load in the next generation of infection.

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 Pillar 1 and 2 case data geolocated to LAD level between 12th June 2020 and 3rd September 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 4th September 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 futher 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.