OPEN ACCESS

Geo-social gradients in predicted COVID-19 prevalence in Great Britain: results from 1 960 242 users of the COVID-19 Symptoms Study app

Ruth C E Bowyer,[1] Thomas Varsavsky,[2] Ellen J Thompson,[1] Carole H Sudre,[2,3] Benjamin A K Murray,[2] Maxim B Freidin,[1] Darioush Yarand,[1] Sajaysurya Ganesh,[4] Joan Capdevila,[4] Elco Bakker,[4] M Jorge Cardoso,[2] Richard Davies ,[4] Jonathan Wolf,[4] Tim D Spector,[1] Sebastien Ourselin,[2] Claire J Steves,[1] Cristina Menni [1]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2020-215119).

[1]Twin Research, King's College London, London, UK
[2]School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK
[3]MRC Unit for Lifelong Health and Ageing, University College London, London, UK
[4]Zoe Global Limited, London, UK

Correspondence to
Dr Cristina Menni, Twin Research, King's College London, London, UK; cristina.menni@kcl.ac.uk

RCEB and TV contributed equally.
CJS and CM contributed equally.

Received 24 April 2020
Revised 23 November 2020
Accepted 24 November 2020

To cite: Bowyer RCE, Varsavsky T, Thompson EJ, *et al*. *Thorax* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/thoraxjnl-2020-215119

ABSTRACT

Understanding the geographical distribution of COVID-19 through the general population is key to the provision of adequate healthcare services. Using self-reported data from 1 960 242 unique users in Great Britain (GB) of the *COVID-19 Symptom Study app*, we estimated that, concurrent to the GB government sanctioning lockdown, COVID-19 was distributed across GB, with evidence of 'urban hotspots'. We found a geo-social gradient associated with predicted disease prevalence suggesting urban areas and areas of higher deprivation are most affected. Our results demonstrate use of self-reported symptoms data to provide focus on geographical areas with identified risk factors.

The COVID-19 epidemic has led to large-scale closures and lockdown measures worldwide with the British government sanctioning lockdown from 23 March 2020 (https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020).

Early in the pandemic, case distribution was not evenly spread across countries, with dense urban centres being the most affected.[1] Individuals in deprived areas have lower life expectancy,[2] are more likely to have multiple underlying comorbidities, have a higher level of influenza-associated hospitalisation[3] and therefore could be more susceptible to COVID-19.[2]

Based on the known socioeconomic health gradient, we hypothesised that individuals in deprived areas were at greater risk of contracting COVID-19. Understanding the geographical distribution of the virus in a socioeconomic context is key to assist adequate healthcare resourcing, particularly intensive care beds.[4]

Here we investigated the geographical distribution of COVID-19 in Great Britain (GB) and its association with area-level deprivation using self-reported data from almost 2 million users of the *COVID-19 Symptom Study*.[5]

We studied 1 960 242 unique GB app users (20–69 years old) reporting on COVID-19 symptoms, hospitalisation, reverse-transcription PCR (RT-PCR) test outcomes, demographic information and pre-existing medical conditions (online supplemental methods) over 23 days (29 March–19 April) of major social distancing measures ('lockdown'). We computed a proxy of contracting COVID-19, based on reported symptoms[6] (positive predicted value=0.69 (0.66; 0.71) (online supplemental methods). We then calculated a predicted prevalence as the proportion of app users that we predicted to have COVID-19 within each area (online supplementary figure S1).

Following aggregation of variables to local authority district level (LAD/geographic unit representing ~17 000 individuals), we tested the geographical distribution of predicted prevalence at eight different time points spanning 23 days. We used Local Moran's I tests, which assess for non-random spatial distribution and clustering of a feature and can be used to identify disease hotspots and cold spots relative to the mean GB predicted prevalence[7] (online supplemental methods).

Next, we used data from the eight different time points and used multivariable mixed-effects models to investigate the association of predicted area-level prevalence (at middle super output area level (MSOA)) and deprivation (as captured by the Index of Multiple Deprivatio) adjusting for different factors including geo-social mediators and confounders (air pollution, general practitioners per MSOA, household density and urbanicity) area level aggregates of obesity and comorbidities) and area-level adjusted mean age and sex and spatial autocorrelations[8] (online supplemental methods).

table table 1 1 and online supplemental table S1. The number of predicted COVID-19 positive individuals ranged between 15 991 and 79 378.

Local Moran's I showed that predicted COVID-19 prevalence clustered in urban areas across GB when considered as a proportion of the population per LAD[7] (figure 1 and online supplemental figure S2) adjusting for multiple testing. Predicted prevalence decreased over time, consistent with 'lockdown' (figure 1 and online supplemental figure S2) (pairwise Wilcoxon rank-sum tests, prevalence: all time points except T2:T3 and T1:T4, $p<0.001$), but some hotspots remained.

In the MSOA-level analysis, area-level deprivation was significantly associated with predicted area-level prevalence in all models (M1–M6, see online supplemental table S2), including in the full model (M6) when adjusting for all geo-social covariates and comorbidities (M6: Beta (95% CI)=−0.15 (−0.17 to −0.130, $p<0.001$). This suggests that

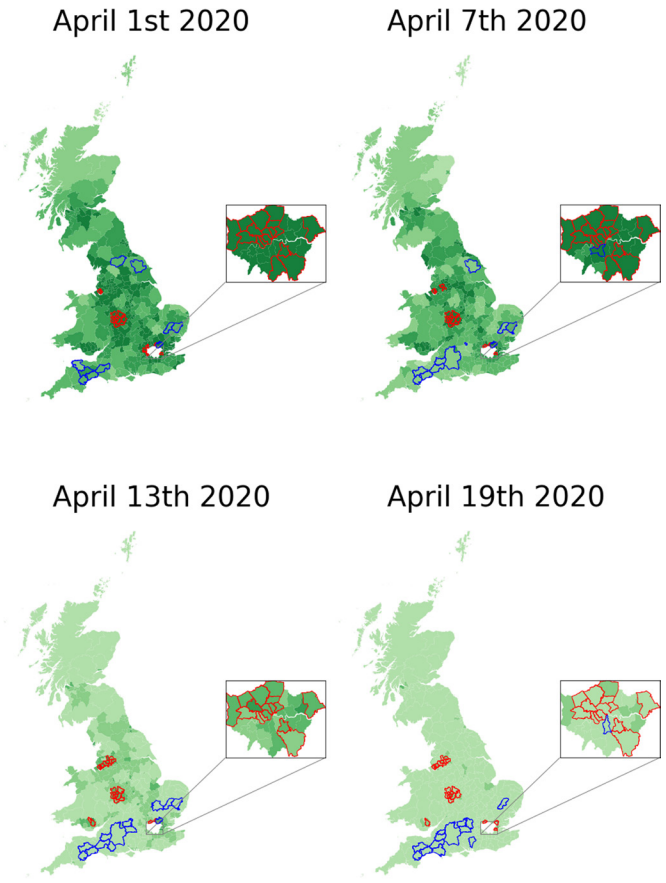


Figure 1 Geographical distribution of predicted COVID-19 prevalence across four time points. Prevalence is presented as proportional to the responders per local authority district (LAD). Analyses are adjusted for multiple testing using Benjamini- Hochberg false discovery rate correction (p<0.05). Inset highlights London where LAD areas are smaller. Hot and cold spots are defined relatively to their neighbours and the mean GB predicted prevalence. Red/blue coloured perimeter lines around each LAD denote hotspot/coldspot.

people in deprived areas were at higher risk.

Predicted COVID-19 prevalence was higher in urban areas compared with rural and in more deprived areas compared with less deprived. This could reflect the likelihood of individuals in more deprived areas working/living with people whose vocations mean they are unable to work from home and are thus more likely to be exposed to circulating COVID-19. Accumulation of socioenvironmental exposures across the life course are known to contribute to a greater health deficit and disease burden[2]; our results suggest that COVID-19 is no exception.

Moreover, our study illustrates how app data could be used to successfully monitor COVID-19 over time and identify hotspots as the viral pandemic progresses and social distancing measures are implemented or eased. Using this method, we detected a geo-social gradient associated with prevalence in the context of COVID-19, suggesting the focus of resources should be on deprived urban areas.

Our study has some limitations and assumptions. We used self-reported data on symptoms that can lead to bias. For example, should users in deprived areas report more symptoms due to a facet of the socioeconomic environment (eg, higher air pollution), this could lead to an incorrectly higher predicted prevalence in deprived areas. Second, app users are a self-selected group, not representative of the general population. Our approach to adjust for age and sex differences at MSOA level is unlikely to sufficiently overcome selection and collider bias.[9] Third, our predicted COVID-19 prevalence is not from confirmed tests via RT-PCR, but rather based on self-reported symptoms. Additionally, we assume that people who have symptoms or have been exposed to COVID-19 are equally likely to use the app as those who do not. We performed a sensitivity analysis by rerunning the pooled analysis on individuals who were self-reportedly healthy at sign up and found the observed associations remained (online supplemental table S3), suggesting selection bias associated with being unhealthy at sign up is not influencing the observed associations of COVID-19 and deprivation. We also assume that people report symptoms in the same way and that their drop-out patterns do not differ by space, time and symptom reports. Finally, we aggregated data at MSOA level that could lead to ecological bias. We also

Table 1 Demographic characteristics of the study population at eight time points

	29 March 2020	1 April 2020	4 April 2020	7 April 2020	10 April 2020	13 April 2020	16 April 2020	19 April 2020	All unique users
N	1 324 843	1 431 515	1 142 923	1 083 601	995 157	985 860	980 608	1 164 262	1 960 242
Predicted COVID-19 (n/%)	60 827	79 378	62 508	48 418	30 132	22 352	16 586	15 991	117 614
	(4.6)	(5.6)	(5.5)	(4.5)	(3.0)	(2.3)	(1.7)	(1.4)	(6.0)
Average number of reports per user	2.9	3.8	4.2	4.7	5	5	5	4.5	4.4
Age, years (median (IQR))	41 (21)	41 (21)	43 (21)	44 (22)	45 (21)	45 (21)	46 (21)	45 (21)	42.2 (21.8)
Male, (n/%)	426 923	459 620	365 078	353 233	327 608	327 620	327 114	388 378	654 950
	(32.2)	(32.1)	(31.9)	(32.6)	(32.9)	(33.3)	(33.3)	(33.4)	(33.4)
Obesity, %	21.3	21.4	20.7	20.3	21.6	22.1	21.4	21.7	21.5
Kidney disease, %	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.5
Lung disease, %	12.2	12.3	12.5	12.5	12.4	12.4	12.4	12.4	12.2
Diabetes, %	2.4	2.5	2.7	2.7	2.8	2.9	2.9	2.9	2.4
Smokers, %	10.5	10.5	9.7	9.4	9.0	8.8	8.7	9.0	10.4
Heartdisease, %	1.4	1.4	1.6	1.6	1.7	1.7	1.7	1.7	1.4

Obesity: BMI >=30 kg/m^2.
At each time point, we only include users who have made an assessment in the previous 7 days. Exclusion criteria are listed in the supplements. Users are asked daily whether (or not) they have any symptoms. Predicted COVID-19 was calculated on users who reported on symptoms. Users who reported having no symptoms were included in the area-level predicted prevalence estimates (please see the supplements for details).
BMI, body mass index.

cannot conclude that deprivation increased COVID-19 prevalence, as there could be unmeasured confounders or other factors.

Future work should check our assumptions and seek to integrate these data with data on area-level morbidity, extended pollution data, ethnicity and disease severity. Indeed, higher mortality has been observed among minority ethnic groups,[10] and disentangling the environmental and biological factors contributing to greater disease burden in both deprived areas and among ethnic minorities is an essential focus of future work to ensure resources and intervention are better assigned.

Twitter M Jorge Cardoso @mjorgecardoso

Acknowledgements We express our sincere thanks to all the participants of the COVID Symptom Study app. We would like to thank the staff of Zoe Global Limited, the Department of Twin Research for their tireless work in contributing to the running of the study and data collection. Finally, we would like to thank Professor Kate Tilling of the University of Bristol for her invaluable insight and help in refining the manuscript.

Contributors Conceived and designed the experiments: CJS, TDS, SO and CM; analysed the data: RCB and TV. Contributed reagents/materials/analysis tools: MF, CHS, BM, MF, DY, SG, JC, ET, EB, MJC, RD and JW wrote the manuscript: RCB, TV and CM; revised the manuscript: all.

Funding Zoe provided in kind support for all aspects of building, running and supporting the app and service to all users worldwide. The Department of Twin Research is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. CM is funded by the Chronic Disease Research Foundation and by the MRC Aim-Hy project grant. CHS is an Alzheimer's Society Junior Fellowship AS-JF-17-011; SO and MJC are funded by the Wellcome/EPSRC Centre for Medical Engineering (WT203148/Z/16/Z), Wellcome Flagship Programme (WT213038/Z/18/Z).

Competing interests TDS is a consultant to Zoe Global Ltd ('Zoe'). SG, JC, EB, RD and JW are or have been employees of Zoe Global Limited. Other authors have no conflict of interest to declare.

Patient consent for publication Not required.

Ethics approval The Ethics for the app has been approved by King's College London ethics Committee (REMAS ID 18210, review reference LRS-19/20-18210), and all users provided consent for non-commercial use. An informal consultation with TwinsUK members over email and social media prior to the app having been launched found that they were overwhelmingly supportive of the project.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Richard Davies http://orcid.org/0000-0003-2050-3994
Cristina Menni http://orcid.org/0000-0001-9790-0571

REFERENCES

1 Stier A, Berman M, Bettencourt L. COVID-19 attack rate increases with City size, 2020. Available: https://papersssrncom/sol3/paperscfm?abstract_id=3564464
2 Marmot M. Health equity in England: the Marmot review 10 years on. *BMJ* 2020;368:m693.
3 Hungerford D, Ibarz-Pavon A, Cleary P, *et al*. Influenza-Associated hospitalisation, vaccine uptake and socioeconomic deprivation in an English City region: an ecological study. *BMJ Open* 2018;8:e023275.
4 Blumenshine P, Reingold A, Egerter S, *et al*. Pandemic influenza planning in the United States from a health disparities perspective. *Emerg Infect Dis* 2008;14:709–15.
5 Drew DA, Nguyen LH, Steves CJ, *et al*. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020;368:1362–7.
6 Menni C, Valdes AM, Freidin MB, *et al*. Real-Time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26:1037–40.
7 Zhang C, Luo L, Xu W, *et al*. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Sci Total Environ* 2008;398:212–21.
8 Anselin L, Griffith DA. Do spatial effects really matter in regression analysis? Papers - Regional Science Association 1988.
9 Griffith GJ, Morris TT, Tudball MJ, *et al*. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020;11:5749.
10 Khunti K, Singh AK, Pareek M, *et al*. Is ethnicity linked to incidence or outcomes of covid-19? *BMJ* 2020;369:m1548.

Supplementary Methods

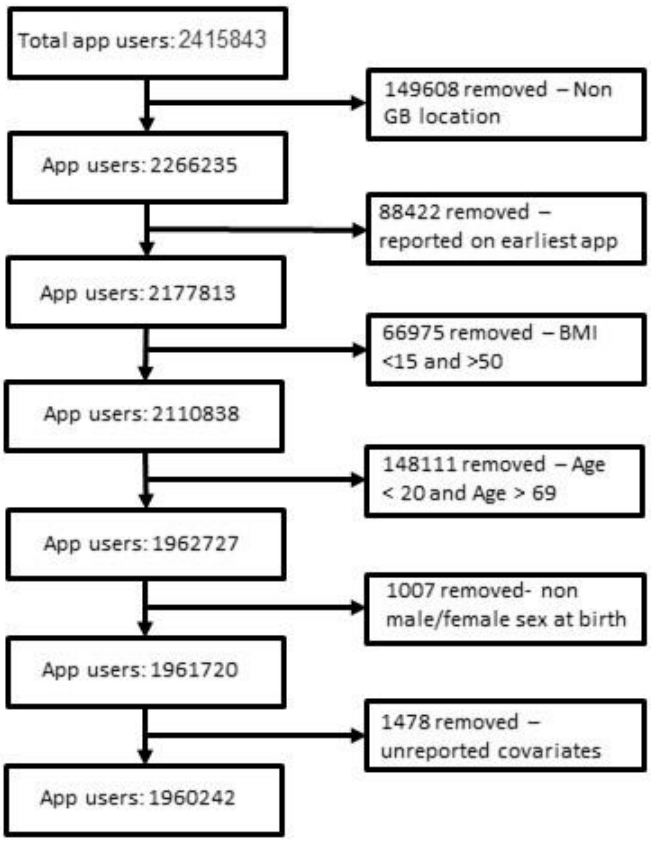
Study setting and participants

The COVID Symptom Study app developed by Zoe with scientific input from researchers and clinicians at King's College London and Massachusetts General Hospital, was launched in GB on Tuesday the 24th March 2020 (https://covid.joinzoe.com/) and in the 23 days (March 29th – April 19th) immediately after the UK lockdown (https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-22-march-2020) was introduced, it reached 2,266,235 unique GB users, making 9,108,769 assessments (e.g. an average user is included in 4 out of 8 timepoints). Referrals/word of mouth, press and eventually partnerships with charities and the Welsh and Scottish governments drove usage.

The app enables capture of self-reported information related to COVID-19 infections. On first use, the app records self-reported location, age, and core health risk factors. With continued use, participants provide daily updates on symptoms, health care visits, COVID-19 testing results, and if they are self-quarantining or seeking health care, including the level of intervention and related outcomes. Individuals without apparent symptoms are also encouraged to use the app. Through direct updates, the research team can add or modify questions in real-time to capture new data to test emerging hypotheses about COVID-19 symptoms and treatments. Importantly, participants enrolled in ongoing epidemiologic studies, clinical cohorts, or clinical trials, can provide informed consent to link data collected through the app in a Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR)-compliant manner with extant study data they have previously provided or may provide in the future.

In this study, we included 1,960,242 unique users as outlined in the flow diagram below (**Figure A**)**.** Briefly, out of 2,415,843 unique app uses who reported on the COVID-19 symptom Study App between 29th March 2020 and 19th April 2020, we excluded (i) 149,608 non GB users; (ii) 88,422 users who only reported on the earliest app-version that did not include loss of smell and taste (the strongest single predictor of COVID-19[1 2]); (iii) 66,975 reporting BMI outside the biological range; (iv) 148,111 users younger than 20 or older than 69; (v) 1007 with missing biological sex at birth or who were not assigned male or female as their biological sex at birth; (v) 1478 users who did not report on pre-existing medical conditions (**Figure A**).

Figure A. Flow diagram representing the study subjects' inclusion criteria.



Geographic clustering of COVID-19 prevalence

Because we were primarily interested in understanding the geography of COVID-19 distribution, and how aspects of an area, in particular area-level deprivation, associated with COVID-19 prevalence we aggregated user data at different GB geographic areas. This was particularly of use as the geosocial variables considered (please see below) are also defined geographically and are time invariant (as they are not defined by the app users themselves but by GB geographic area).

The maps (**Figure 1, S2**) were created using a shapefile of Local Authority Districts (LADs) from the Office for National Statistics (ONS) using the geopandas package in Python. Overlaid on the map are statistically significant 'hot-spots' and 'cold-spots' at LAD level. To assess the significance of these regions, we used Local Moran's I test, as introduced below. In order to do this, spatial weights were

calculated to create a spatially lagged COVID-19 prevalence variable for each LAD. Because our geographical units share borders we assume a queen criterion, which assumes equal weights of neighbouring areas, which is appropriate for defining these. Islands were considered to have zero neighbours. We adjusted for multiple testing using the Benjamini & Hochberg method ('p.adjust') and used the 'spdep' package in R for the Local Moran's and calculation of the spatial lag. This approach of calculating the spatial lag was repeated at the middle super output area level (MSOA) level (below).

Hotspot and Coldspot definition

Predicted prevalence hotspots at LAD levels were defined using Local Moran' s I. The Moran's I statistic gives a value indicating the spatial clustering of a variable relative to its neighbours. Where there are significant (false discovery rate (FDR)adjusted $p < 0.05$) high positive local Moran's I in high value neighbourhood (i.e. where the significant area also had a predicted prevalence greater than the mean predicted prevalence and greater than the mean of the lagged variable, which effectively represents how similar COVID-19 prevalence is to the areas that surround it) this implies the area can be considered a 'hotspot'[3]. This method ensures we do not consider areas as hotspots where they may have higher predicted prevalence to the surrounding areas but are lower than average for the UK, although it might miss areas that are surrounded on all borders by other areas which would be considered hotspots. A coldspot is assessed similarly using Local Moran's I, but where the area is less than the mean and mean of the lagged variable.

Sources of geographic data

Index of Multiple Deprivation (IMD)

The IMD was downloaded from the relevant government websites as below, and the most recent IMD available at time of analysis was used:

- English (2019): https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019
- Scottish (2016): https://www2.gov.scot/Topics/Statistics/SIMD
- Welsh (2019): https://statswales.gov.wales/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/WIMD-2019

Because the IMD is calculated in each devolved administration using slightly different methodology, and because of the different number of areas in each country, ranks are not directly comparable.

Bowyer RCE, *et al. Thorax* 2020;0:1–3. doi: 10.1136/thoraxjnl-2020-215119

Therefore, we used within-country defined deciles. As the IMD is calculated for smaller area geographies than MSOA, we calculated the average IMD per MSOA. This was then categorised into quintiles where 1 is the least deprived and 5 is the most deprived.

Rural-urban gradient (RUC)

The RUC was downloaded from the relevant government websites as below:

- England and Wales RUC (2011): https://data.gov.uk/dataset/9c0e093d-d267-4eb8-90d8-54475ab4d1ff/rural-urban-classification-2011-of-middle-layer-super-output-areas-in-england-and-wales
- Scotland RUC (8 fold classification): https://www2.gov.scot/Topics/Statistics/About/Methodology/UrbanRuralClassification

The resulting scale runs from 1 – 8, where 1 is the most urban and 8 is the least.

Nitrogen Oxide (NOx) data

We used NOx pollution data from the Department of Environment, Food and Rural Affairs (https://uk-air.defra.gov.uk/data/) for England, Scotland and Wales from 2018. Data is provided with Ordinance Survey 1km^2 grid resolution which was used to calculate per MSOA air pollution by taking the area-weighted average of the readings.

General Practitioners (GPs)/MSOA

GPs addresses were used to derive the number of GPs from each MSOA, from the following data sources:

- England & Wales:https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data
- Scotland: https://www.opendata.nhs.scot/ne/dataset/general-practitioner-contact-details/resource/b092b69f-0838-408e-bb89-082562f0e1cd

Average household number

This figure was derived from data by dividing the number of houses with at least one usual occupant with the total population for the same area.

Data sources for occupancy data were downloaded from the following sources:

- England & Wales (table PHP01 2011): https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/households/household-estimates/small-area-statistics-on-households-and-dwellings
- Scotland: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/households/household-estimates/small-area-statistics-on-households-and-dwellings

MSOA-level mixed-effects models

We employed multivariable mixed-effects models to understand the relationship of predicted COVID-19 prevalence at MSOA level with deprivation. As a reminder, these models were ran at MSOA-level rather than individual-level. This included the following variables:

The Index of Multiple Deprivation, our primary explanatory variable (IMD, categorised into quintiles generated on the average IMD within each MSOA, where 1 is most deprived and 5 is least, and considered as a continuous variable).

Other considered geosocial factors included a rural-urban gradient (RUC, considered as a continuous variable where 1 is the most urban and 8 is the most rural), General practitioners per population in MSOA (GPs/MSOA, where a higher number indicates more GPs per individual by MSOA), average household number (calculated as number of inhabited dwellings/MSOA population, where a higher number indicates a higher average number of individuals per household). Because it was on a very different scale to the rest of the predictor variables, GPs/MSOA was scaled to have mean 0 and 1 SD prior to model inclusion.

We additionally adjusted for the following variables derived from app response data, considered as percentage of responders within the MSOA: those who reported having kidney, heart or lung disease, and who are diabetic, a smoker or obese (calculated as BMI<30). We derived mean-adjusted age and sex variables to partially adjust for response bias (i.e. the extent responders in an MSOA represented the demographic of that MSOA). This was calculated as the difference of the expected mean/ratio of age/sex in the MSOA (derived from ONS population data) and the observed mean/ratio of age/sex amongst respondents.

We included a spatial lagged variable of the COVID-19 prevalence outcome. Inclusion of the lagged variable is one method that accounts for spatial autocorrelation (SAC)[4]. It attempts to adjust for spatial autocorrelation by capturing the variance explained by the influence of neighbouring regions on the value of interest – in this case COVID-19 severity/prevalence. The lagged variable is calculated at MSOA level by applying a spatial weights matrix (calculated in this instance under queen's

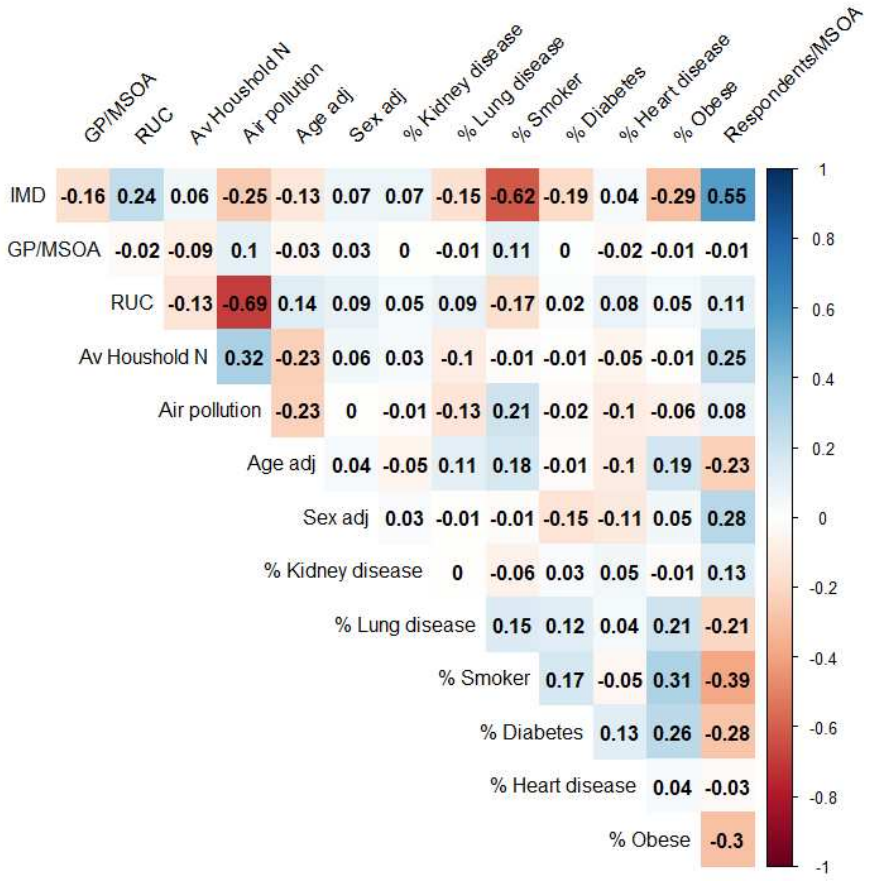
contiguity) to the outcome variable (in this case COVID-19 prevalence) and computing the lag using the function lag.listw in the 'spdep' R package. This variable is then included as a covariate within the model.

Data from eight time points were analysed , calculating the covariates (derived from app responders) and spatial lag at each time point, a dummy variable adjusting for the different sample times was included in the model as a random effect (allowing for a random intercept). MSOA was also included to allow for a random intercept to account for the repeat observations over the eight time periods, along with country as a fixed effect to account for difference in methodology in creation of IMD and RUC.

The users' distribution across GB is not uniform but all analyses took this into account by considering only middle super output areas (MSOAs) with at least 20 individuals reporting on the app (n = 8097, n removed = 387), and we included as a covariate the proportion of responders per MSOA at each time point, in order to adjust for differences in responders by MSOA. Analysis was conducted in RStudio v1.1.423 and R v3.6.3.

Variables were checked for multicollinearity before model inclusion using Spearman's correlation, (see **Figure B**) with the *a priori* threshold of > (+/-) 0.7 indicating a variable should be removed.

Figure B. Assessment of collinearity between the variables included in the MSOA-level mixed-effects models. Each cell of the matrix displays Spearman's correlation between two. The table is colour coded according to the Spearman's correlation, with blue denoting a positive correlation and red denoting a negative correlation. GP/MSOA= General Practitioners per middle super output area level; RUC= Rural-urban gradient; Av Household N= average household number.



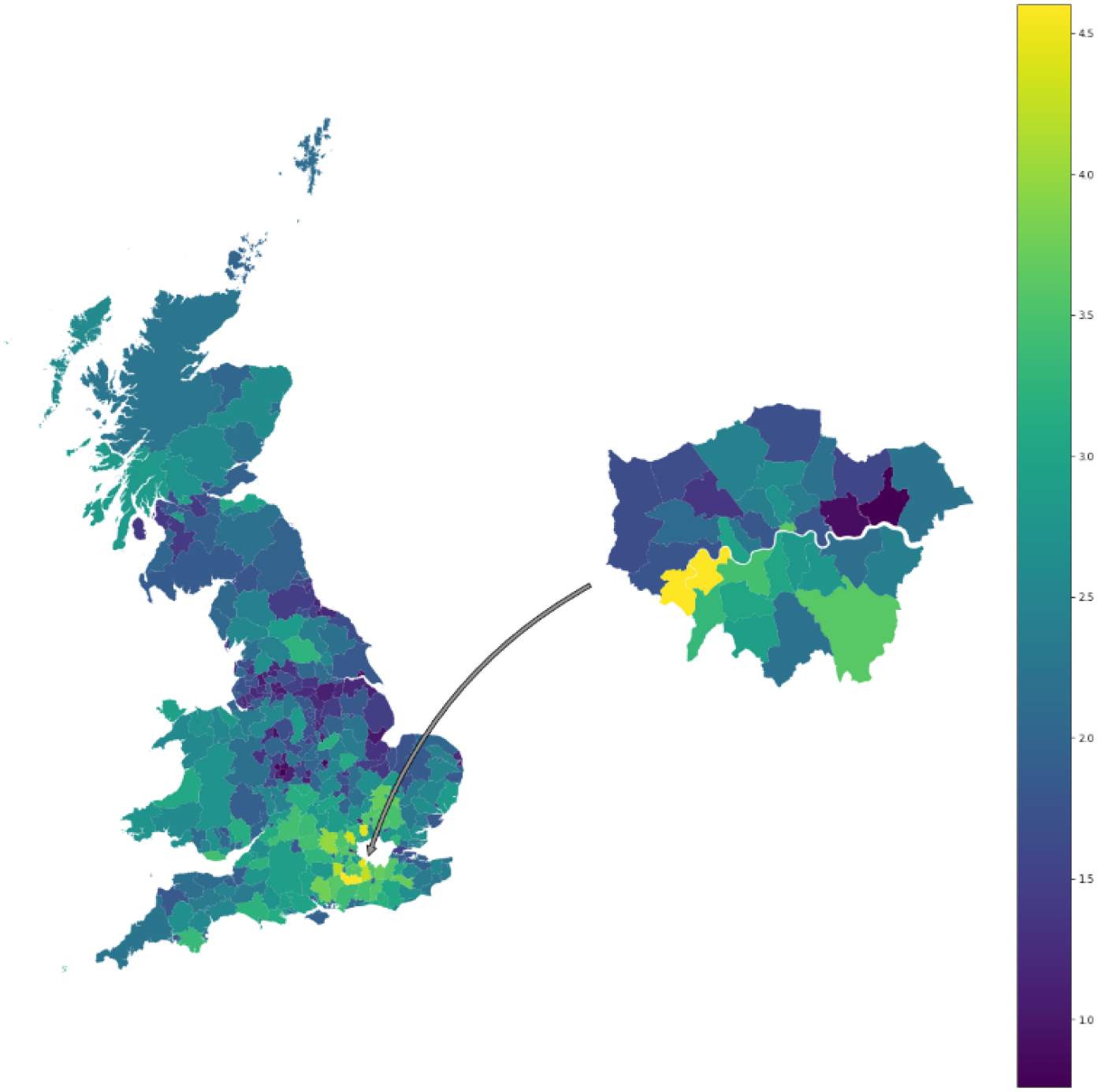
The model approach was therefore as follows:

- Model 1 (M1): Linear regression of the estimated COVID-19 prevalence and the IMD
- Model 2 (M2): Linear mixed effects model (LMM) of estimated COVID-19 prevalence and the IMD, adjusted for country, and allowed a random effect of MSOA ID and time (assuming random intercept for both)
- Model 3 (M3): Linear mixed effects model of estimated COVID-19 prevalence and the IMD, adjusted as above in M2, with additional adjustment for spatial autocorrelation (SAC) via inclusion of a spatial lag.

- Model 4 (M4): Linear mixed effects model as in M3, with the inclusion of geosocial mediators and confounders and proportion of MSOA population who were app users.
- Model 5 (M5): Linear mixed effects model as in M4, with the inclusion of aggregated co-morbidities as the % of respondents in MSOA with diabetes, kidney, lung or heart disease, who are obese or are smokers.
- Model 6 (M6): Covariate + mean-adjusted LMM – Linear mixed effects model as in M6, with the inclusion of mean-adjusted age and sex variables

Supplementary References

1. Menni C, Sudre CH, Steves CJ, et al. Quantifying additional COVID-19 symptoms will save lives. *Lancet* 2020;395(10241):e107-e08. doi: 10.1016/s0140-6736(20)31281-2 [published Online First: 2020/06/09]
2. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;26(7):1037-40. doi: 10.1038/s41591-020-0916-2 [published Online First: 2020/05/13]
3. Zhang C, Luo L, Xu W, et al. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. *Sci Total Environ* 2008;398(1-3):212-21. doi: 10.1016/j.scitotenv.2008.03.011 [published Online First: 2008/04/29]
4. Diniz-Filho JA, Nabout JC, de Campos Telles MP, et al. A review of techniques for spatial modeling in geographical, conservation and landscape genetics. *Genet Mol Biol* 2009;32(2):203-11. doi: 10.1590/S1415-47572009000200001 [published Online First: 2009/04/01]


4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0

Predicted Covid-19 +ve cases in GB with highlighted spatially significant hotspots