Model-based evaluation of the role of deprivation on deaths due to COVID-19: A summary report for the Scottish Coronavirus Model (SCoVMod)

Authors:

Banks C.J.*[1], Colman E.*[1], Doherty T.* [1], Tearne O.*[2], Arnold M.[2], Balaz D.[1], Beaunée G.[3], Bessell P.[1], Enright J.[4], Kleczkowski A.[5], Rossi G. [1], Ruget A.-S. [1], Kao R.R.† [1,6]

* These authors contributed equally to this work.
†Author for correspondence.

Summary.

Deprivation is recognised as an important factor in determining the likelihood of mortality following infection with COVID-19, with several studies pointing to both demographic and health-related indices as having a role in placing a higher burden on deprived areas. However, disentangling factors related to transmission dynamics (influencing who becomes infected) and those related to health (increased mortality once infection has occurred) is difficult, due to uncertainty regarding the underlying infection rates, and strong associations across multiple indices that potentially influence both. Here, we use an individual-based model of COVID-19 to evaluate the role of deprivation in the Scotland COVID-19 epidemic, where data on deprivation are available at a high geographical resolution.

The Scotland Coronavirus Model (SCoVMod) is an agent-based simulation of COVID-19 spread that utilises descriptions of individual-level activity to propagate infection across a virtual national landscape. At it operates at the individual level with individual level movement, it directly accounts for phased changes in restrictions on movement, or the national level implications of changes in the person-to-person transmission rate. In this analysis, we consider the early epidemic in Scotland, fitting the model in the period from the week of the first COVID-19 related death (week ending March 15th) up to the week when lockdown restrictions were observed to have a substantial impact on disease reproduction numbers (week ending April 5th). We infer in the model the dependence of COVID-19 mortality rates on mean "Health Index" per Health Board (of which there are 14 in Scotland). We show that introducing reductions in commuter mobility consistent with published reports, and imposing a reduction in the reproduction rate consistent with observed estimates post April 5th, creates a trajectory consistent with the observed data until mid-May. Fitted cumulative mortality across Health Boards lie within 95% of simulations in all Health Boards where substantial infections occur. The best fit model shows a considerable influence of health index on mortality; the most likely impact is that the Health Board with the best index will have, on average, a reduction to approximately 30% of the mortality compared to the Health Board with the lowest health index. These results should aid in planning for possible future outbreaks, either as we move out of lockdown, or in a putative future COVID-19 epidemic.

[1] Roslin Institute, University of Edinburgh
[2] Animal and Plant Health Agency, Weybridge.
[3] INRAE, Nantes, France.
[4] School of Computing Science, University of Glasgow
[5] Mathematics & Statistics, University of Strathclyde
[6] Royal (Dick) School of Veterinary Studies.

1. Model overview/methods.

A schematic overview of the model is presented in figure 1. Further details are provided in the following section.

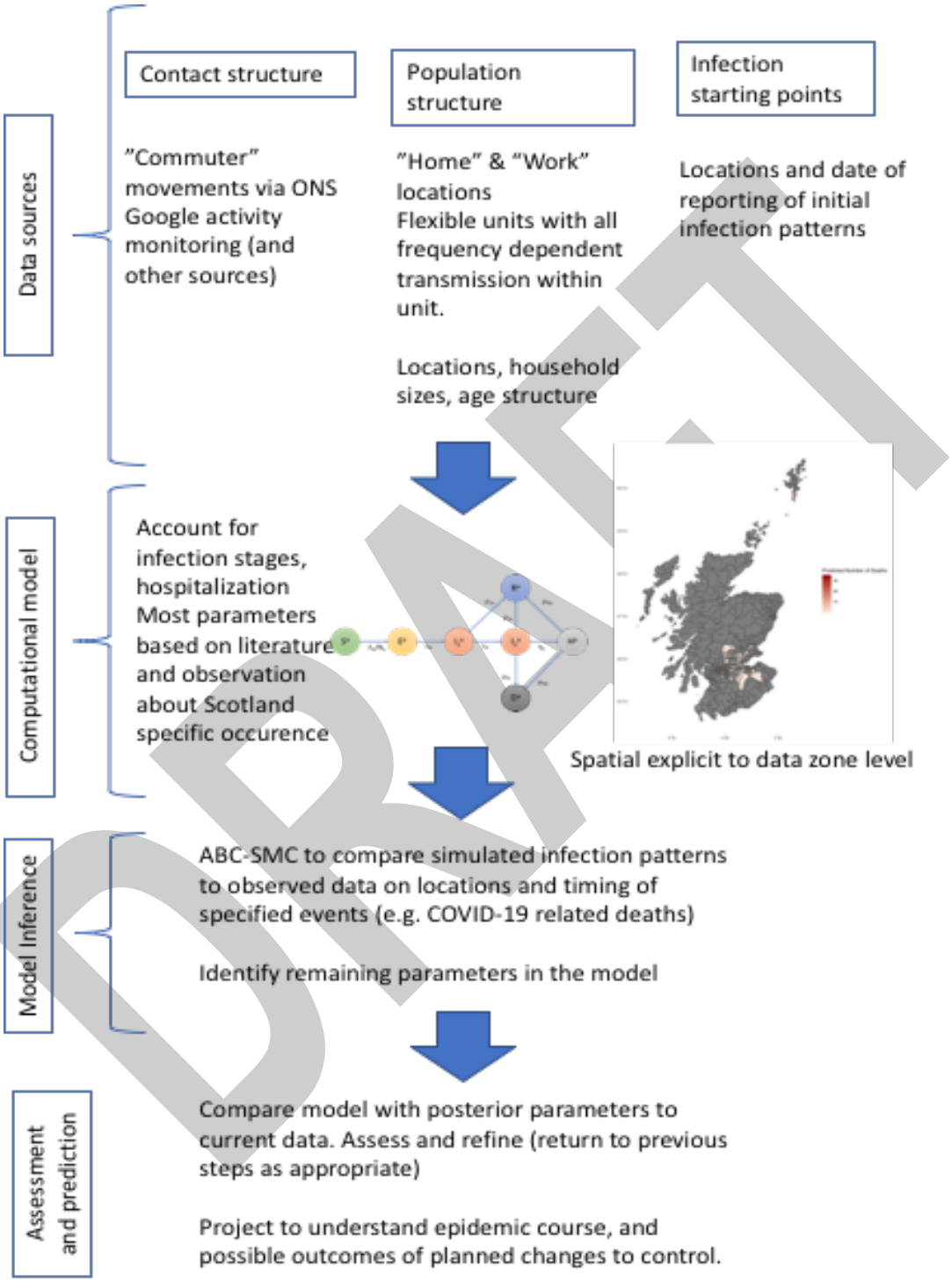

Figure 1. Overview of SCoVMod, including data assimilation, model development, model assessment and prediction.

1.i. Compartmental model.

SCoVMod considers key aspects of COVID-19 epidemiology, including a latent phase, mildly infectious and highly infectious individuals, hospitalisation, recovery and death, similar to models used for other investigations[7,8]. These epidemiological processes are embedded in a compartmental model, as described in the schematic of Figure 1. Within population units (in this case, Census output areas or OAs[9,10]) the model assumes homogeneous mixing. The infection model is embedded into a population movement structure which includes patterns of home and work contact. Individuals in infectious stages (and at varying ages) will have the potential to infect others when co-located in the same OA at the same time (considering date and day/night patterns). Infection is then moved spatially according to a network-based model of commuter movement.

We assume that while the death rates for all age classes are the same, recovery rates differ, resulting in age-dependent differences in final outcome (see Table 1 for supporting references). All exposed individuals are assumed to become infectious. It is assumed that the spatial patterns of infection are driven by commuter movements; i.e. transient movements between OAs lasting less than a day. We do not consider overnight shifts in location or introductions from outside Scotland beyond the impact on the initial seeding. Infection dynamics are simulated via a tau-leap algorithm using half day timesteps.[11]

Because our analysis concentrates on the early stages of the epidemic, we make a number of simplifying assumptions regarding transmission pathways: infections in care homes are assumed to result in few additional infections outside of these locations, and are therefore not considered separately, with a similar assumption for hospitals.[12] We assume that only adults contribute to commuter movement; in the daytime, the remaining proportion of adults and all young and elderly individuals are assumed to move primarily within their local OAs, which also account for non-work activities (e.g. interactions in shops). Commuting is restricted to healthy and exposed or mildly infected individuals; severely infected and hospitalised individuals do not commute.

Additional reports show that deprivation is an important indicator of COVID-19 mortality[13]. Therefore in addition, we consider the role of deprivation by adjusting for health in the model. Here, we use these factors to drive regionally specific differences, allowing for health index adjusted mortality rates by health board.

The equations for the complete model are provided in Appendix I.

---

[7] Arenas, A., Cota, W., Gómez-Gardenes, J., Gómez, S., Granell, C., Matamalas, J. T., ... & Steinegger, B. (2020). A mathematical model for the spatiotemporal epidemic spreading of COVID19. medRxiv. https://doi.org/10.1101/2020.03.21.20040022

[8] Di Domenico, L., Pullano, G., Sabbatini, C. E., Boëlle, P. Y., & Colizza, V. (2020). Expected impact of lockdown in Île-de-France and possible exit strategies. medRxiv. https://doi.org/10.1101/2020.04.13.20063933

[9] https://www.scotlandscensus.gov.uk/variables-classification/output-area-2011

[10] https://www.nrscotland.gov.uk/files//geography/2011-census/geography-bckground-info-comparison-of-thresholds.pdf

[11] Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. The Journal of chemical physics, 115(4), 1716-1733. https://doi.org/10.1063/1.1378322

[12] https://www.thelancet.com/lancet/article/s0140-6736(20)31100-4

[13] https://www.medrxiv.org/content/10.1101/2020.05.06.20092999v1.full.pdf

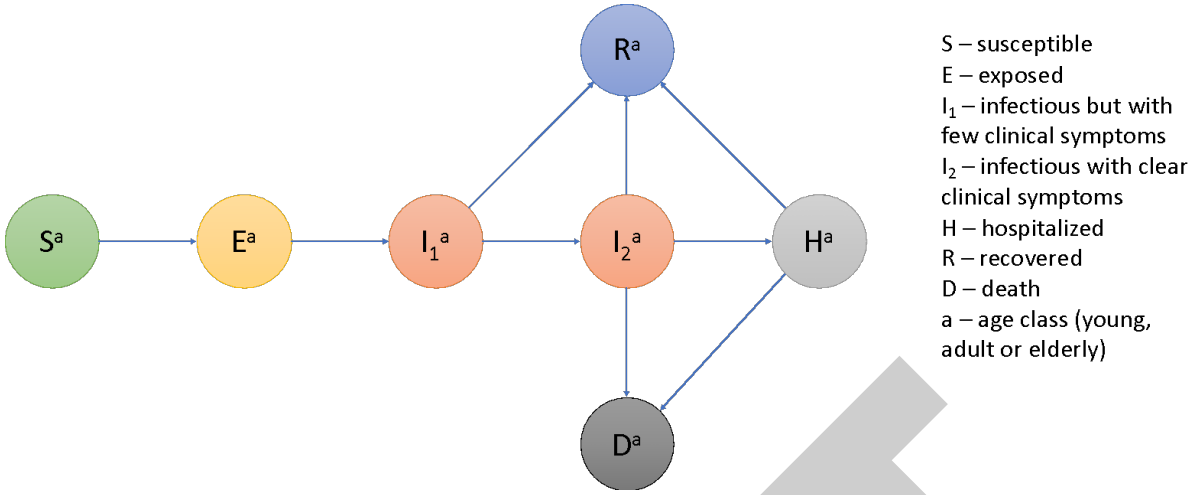

Figure 2. Schematic of infection stages in SCoVMod. Individuals pass through stages post infection as described by arrows. Not all stages are obligatory for all infected individuals (e.g. some individuals recover without going to hospital).

1.ii Data sources and generation of commuter movement patterns.

The model population demographics, including patterns of commuting are all derived from census data. Combining data sources creates a network of interactions of OAs across all of Scotland (see figure 3). Full details of movement pattern generation are found in appendix II.

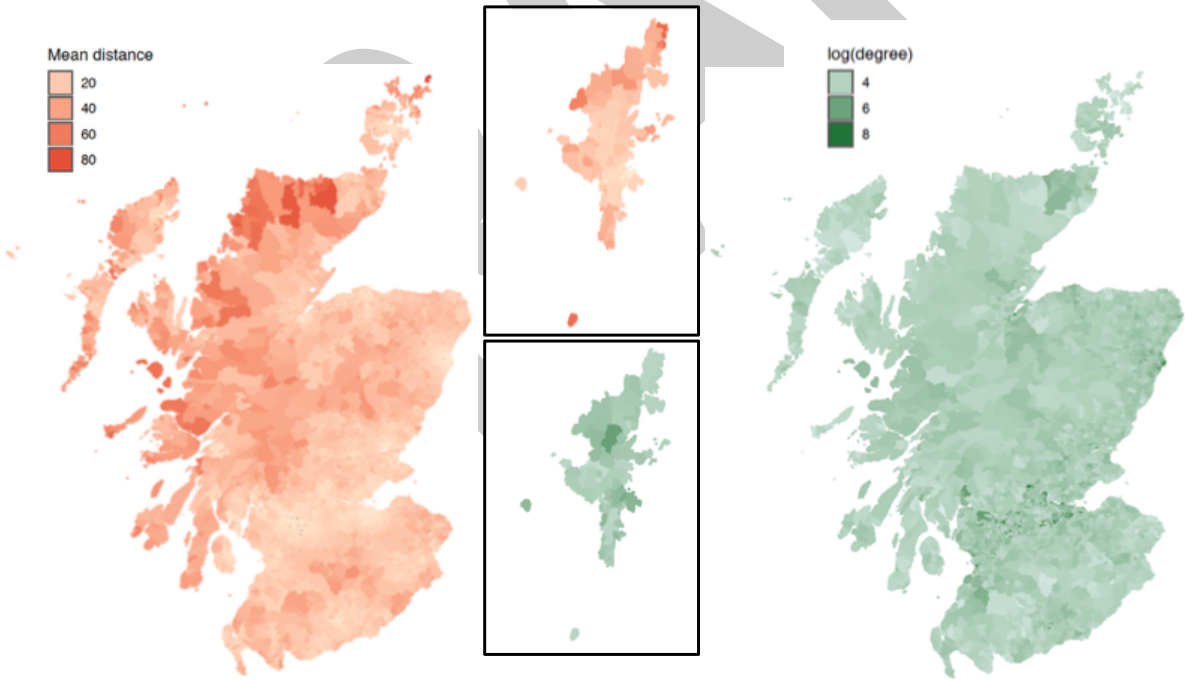

Figure 3. Commuter movement patterns, with commuters aggregated by output area (OA; Census areas with typically 50-500 individuals; maximum of 2081). Data according to the 2011 census. (L) mean distance travelled from OA in km. and (R) mean number of OA's to which each OA is connected to. The greatest distances are travelled on average by individuals in remote locations. The greatest network degree is found in highly urbanised areas.

1.iii Model Inference and Computational Background.

The population structure, movement patterns and the infection model are used to generate simulated epidemics which iterate over half-day increments to simulate two types of contact, non-

commuting/home locations, and locations where commuting individuals interact. Homogenous mixing at the OA level is assumed. Simulated epidemics are compared in space and time to the recorded pattern of COVID-19 spread in Scotland.

All population and statistics on COVID-19 deaths are drawn from National Records Scotland (NRS) using the table of registered deaths.[14] The number and geographical distribution of individuals tested for COVID-19 in Scotland were not available at point of this analysis and was likely to have some biases, particularly in the earlier stages of COVID-19 spread. Also registration of deaths is required within 8 days of the event, introducing an uncertainty in the date of death for each case. In our inference, all non-observable or unknown parameters were therefore estimated using recorded figures for deaths due to COVID-19 related causes, considering all weeks beginning 9th March and ending on the 12th April 2020 and assuming that for each reported case, death occurs in the week prior to the registered week. As an evaluation of the full likelihood evaluation of the model would not be straightforward, estimation was performed using a sequential Monte Carlo implementation of Approximate Bayesian Computation (ABC- SMC).[15,16] Summary statistics used for the inference were defined as the weekly number of deaths due to COVID-19, aggregated at the Health Board level (https://www.scot.nhs.uk/organisations/). The metric used to compare simulated and observed summary statistics was defined as a sum of squared errors (SSE) of the number of deaths due to COVID-19, over the health board, recorded weekly:

$$score = \sum_{all\ weeks} (D_{sim} - D_{obs})^2$$

Where:

D_{sim} = no. dead per health board simulated
D_{obs} = no. dead per health board observed

Uniform prior distributions were applied to parameters to constrain their values to ranges that are plausible based on the available literature relevant to the early, pre-lockdown period (Table 1). Infection statistics are taken from the weekly NRS statistics for deaths registered with COVID-19 related causes[17] – while these are the most complete available, they also differ slightly from other official sources which record the date of death, rather than the date of registration of death.

The ABC-SMC inference requires $5x10^4$ - 10^5 model simulation runs before model convergence. Each simulation takes approximately 2-3 minutes to run. The inference framework is run on a distributed application framework (Akka)[18]. Running Akka on a Cloud Computing IAAS infrastructure (Amazon AWS[19]) allows for rapid scaling upwards to 16Gb and 4 cores per computer node and outwards to 200 compute nodes. In the inference framework each "generation" of the ABC-SMC therefore is complete in approximately 20-60 minutes, with the tolerance in the acceptance scheme usually stabilising after no more than 15 generation (indicating a 'settled' posterior estimate).

---

14 NRS statistics: https://statistics.gov.scot/

15 e.g. Hartig, F., Calabrese, J.M., Reineking, B., Wiegand, T. & Huth, A. (2011) Statistical inference for stochastic simulation models – theory and application. Ecology Letters, 14, 816–827.

16 Toni et al. "Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems". " J. Roy. Soc. Interface https://doi.org/10.1098/rsif.2008.0172

17 https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/vital-events/general-publications/weekly-and-monthly-data-on-births-and-deaths/deaths-involving-coronavirus-covid-19-in-scotland

18 https://akka.io

19 https://aws.amazon.com

The model code has been written using industry grade software engineering practices, i.e. Agile development for project task planning[20], Test Driven Development (TDD)[21], Pair programming and code reviews to produce unit tested, robust and reusable software components. The majority of the code has been reviewed by a second software developer (the remaining review is ongoing).

1.v Model Seeding

Infected seed individuals are distributed according to observed case data. As case data represents chiefly severely symptomatic or hospitalised individuals, we assume those individuals are present in the population in the 5 days prior to the recorded date of observation. Recorded statistics for COVID-19 infections are evaluated here at the Health Board level; while some detail below this scale is available, in order to achieve summary statistics in the early stages, we use this approximate approach. We note that this approximate approach should capture the relative proportion of individuals per health board.

1.vi Modelling Lockdown

We propose that the impact of lockdown in our model can be modelled by a combination of two effects. First, the reduction in longer distance activity requires that we reduce the volume of commuter activity. For this we reduce it in a manner consistent with the observed decline seen in the Google mobility data[22]. We compared this to an independent dataset to corroborate the reduction in rural areas, where data are few (see appendix III). Second, social distancing will introduce a reduction in contact and therefore transmission rates, that will also be impacted by local 'saturation' effects (due to clustering of contacts, such as would occur when all people in a household have greater contacts with each other, than outside). In order to consider this impact of lockdown, we note that the reproduction number of COVID-19 is independently estimated to be high and consistent until approximately April 6th, then drops below 1 with a reduction to approximately 0.30 of the original value [23]; this is also consistent with an estimate of the impact of movement rate reduction in London in late March.[24] We use this to reduce contact rates and therefore the transmission rates to 30% of the original value.

1.vii Scottish Index of Multiple Deprivation (SIMD).

Based on NRS Covid-19 data from week-18 (published 6thMay) complete up to 3rdMay, we examined the impact of different deprivation factors. We found that whilst deprivation overall (as measured by the Scottish Index of Multiple Deprivation) is significantly associated with increased Covid-19 mortality this can be further disaggregated:

1. Population level risk of Covid-19 mortality is associated with the SIMD indicator that describes (good) accessibility and orthogonally with the SIMD indicator that describes

[20] Kent Beck; James Grenning; Robert C. Martin; Mike Beedle; Jim Highsmith; Steve Mellor; Arie van Bennekum; Andrew Hunt; Ken Schwaber; Alistair Cockburn; Ron Jeffries; Jeff Sutherland; Ward Cunningham; Jon Kern; Dave Thomas; Martin Fowler; Brian Marick (2001). "Manifesto for Agile Software Development".

[21] Beck, K., Test-Driven Development By Example, Addison- Wesley, Boston, MA, USA, 2003.

[22] Google Community Mobility Reports: https://www.google.com/covid19/mobility/

[23] Covid-19: Framework for Decision Making, Further Information (published 23rd April, 2020), accessed 11th May, 2020 https://www.gov.scot/publications/coronavirus-covid-19-framework-decision-making-further-information/

[24] https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-020-01597-8

(poor) health. Indicating that areas with poorest health and good access experienced higher Covid-19 mortality.

2. Risk of excess Covid-19 mortality (Covid-19 deaths as a fraction of all deaths) is most closely associated with the access indicator component of SIMD. This indicates that the areas that have good local connectivity and transport will have higher rates of Covid-19 transmission.

We hypothesize that "access" is a proxy for transmission model dynamics in two ways; first, transmission rates in the model may be influenced by access (with greater access likely implying a higher probability of introduction of infection and therefore earlier introduction. Thus the observed differences in mortality overall is dependent on the time since introduction (and therefore the initial seeding plus transmission dynamics), and health. We therefore fit the health index (only) using a simple model:

$$\mu_{hb} = \mu_{av} * b * \left\{1 + \left(\frac{\kappa_{HB} - \kappa_{av}}{\kappa_{av}}\right)\right\}$$

Where μ_{hb} is the COVID-19 related mortality rate for a given health board, μ_{av} Is the average across health boards, κ_{HB} a is the health board mean health index value (from the SIMD), and "b" is a fitted parameter given a prior range of {0.0.75} in order to precludes negative values for low values of κ_{HB}.

1.viii. Model Parameters – Priors for fitted parameters and values for configured parameters.

Epidemiological parameters are based upon the extant literature, with studies that are epidemiological based and GB-centric informally preferred. A list of parameters (6 fitted, 7 configured) in the model and their sources are found below in Table 1.

Table 1. Epidemiological parameters in SCoVMod, with priors and fixed values as appropriate. Where age is not indicated, parameters are assumed not to be age dependent. References for parameters are found in Appendix IV.

Parameters	Transition	Symbol	Age	Central estimate	Prior	References
Latency period (days)	$E \rightarrow I_M$	$1/\gamma$	All	*fitted*	*U(2,28)*	1-5
Time from mild infectiousness to recovery (days)	$I_M \rightarrow R$	$1/\rho_M$	All	*fitted*	*U(2,28)*	2,5
Symptoms onset time after infectiousness (i.e. incubation period − latency period) (days)	$I_M \rightarrow I_S$	$1/\gamma_M$	All	*fitted*	*U(2.5,28)*	1-14
Transmission rate for severe infectors I_S (baseline, daytime)	$S \rightarrow E$	β_d	All	*fitted*	*U(0,2)*	-
Transmission rate for severe infectors I_S (baseline, nightime)	$S \rightarrow E$	β_N	All	*fitted*	*U(0,2)*	-
Transmission rate multiplier for mild infectors I_M	$S \rightarrow E$	y	All	*fitted*	*U(0,1)*	-
Time from symptoms onset to hospitalization (days)	$I_S \rightarrow H$	$1/\eta$	All	4.00	-	6,7,10,12,13, 15-21
Time from symptoms onset to recovery (days)	$I_S \rightarrow R$	$1/\rho_S$	Young Adults Elderly	19.00 20.70 21.60	-	6,16
Time from hospitalization to recovery (days)	$H \rightarrow R$	$1/\rho_H$	All	12.35	-	8, 10,13,16
Time from symptoms onset to death (days)	$I_S \rightarrow D$	$1/\mu S$	All	16.00	-	10,12,15,16, 20,21
Time from hospitalization to death (days)	$H \rightarrow D$	$1/\mu_H$	All	9.70	-	6-8,10,12,16

2. **Results – parameter posterior values. References in table are found in a separate appendix (vi).**

Posterior parameter distributions are shown in figure 3. The univariate posteriors show substantial opinions across all parameters.

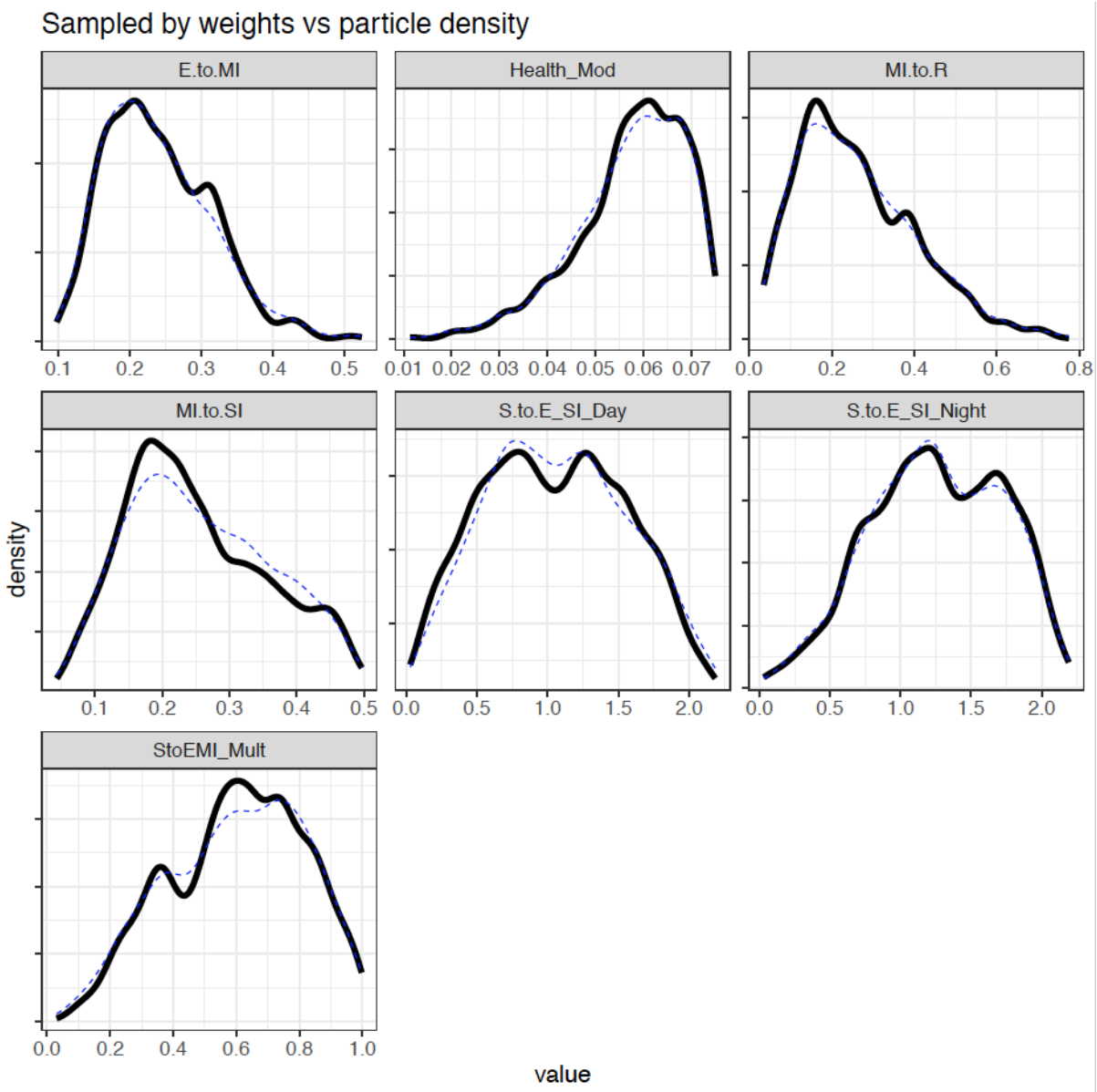

Figure 3. Posterior distributions of fitted parameters, from top L to lower R: (i) transition rate from exposed to mildly infectious (per half day), (ii) health index associate modifier (iii) recovery rate for mildly infectious individuals (per half day), (iv) transition rate from mildly infectious to severely infectious (per half day), (v) frequency dependent transmission rate for severely infectious individuals in "day" locations (per 5 severely infectious individuals, per half day), (vi) frequency dependent transmission rate for severely infectious individuals in "night" locations (per severely infectious, per half day), (vii) multiplier for mildly infectious individuals). Dotted line in each panel shows estimate of the posterior in the penultimate generation.

2.ii Model Fit

The fitted model is shown to reproduce the epidemic curve over the fitted period across most health boards (Figure 4) with good fidelity, with the exceptions being the Highlands (excess mortality over predicted) and Shetland (less mortality than expected in the model).

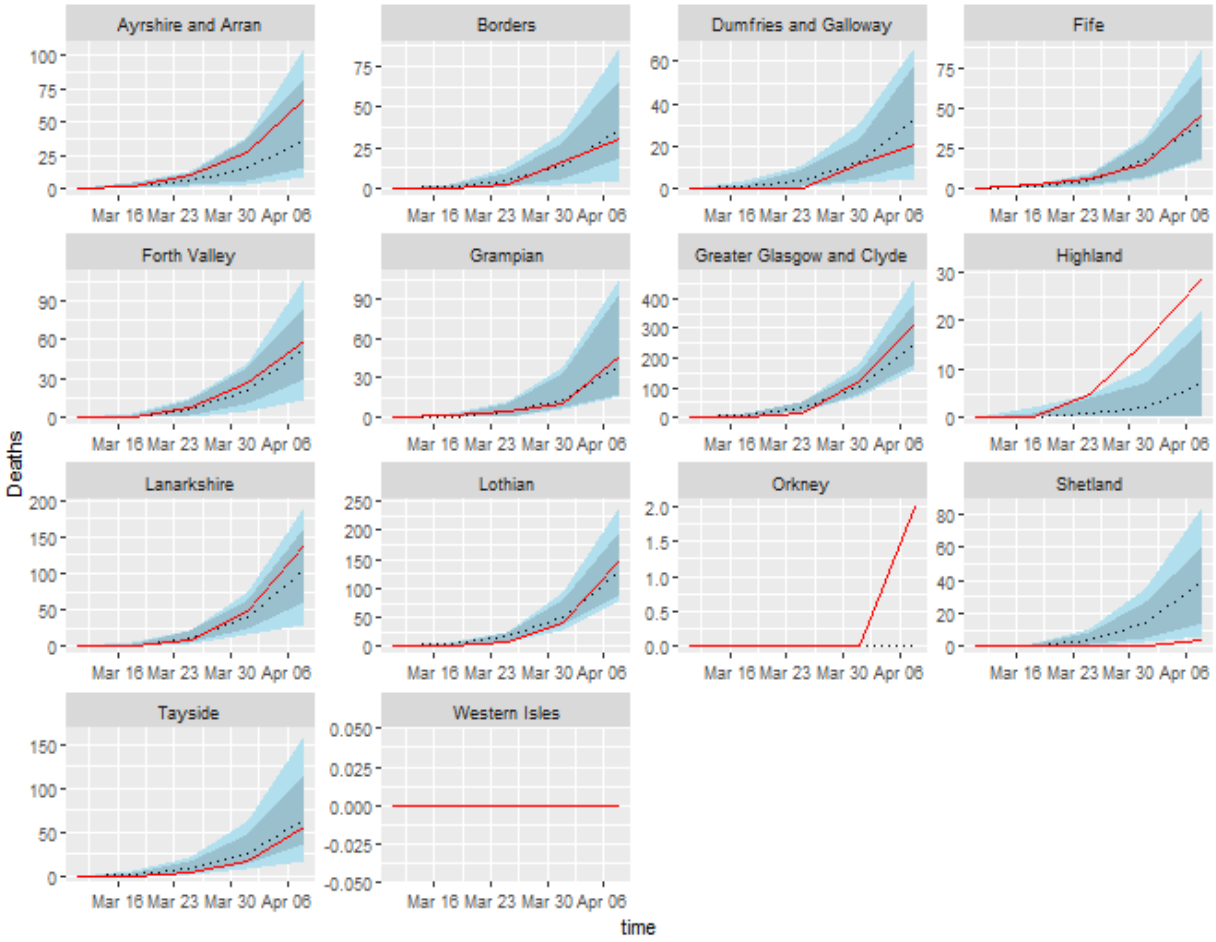

Figure 4. Registered number of COVID-19 related deaths (red), median of 50 simulations (dotted line), and 80% and 95% C.I.'s (dark and light blue respectively), aggregated by Health Board.

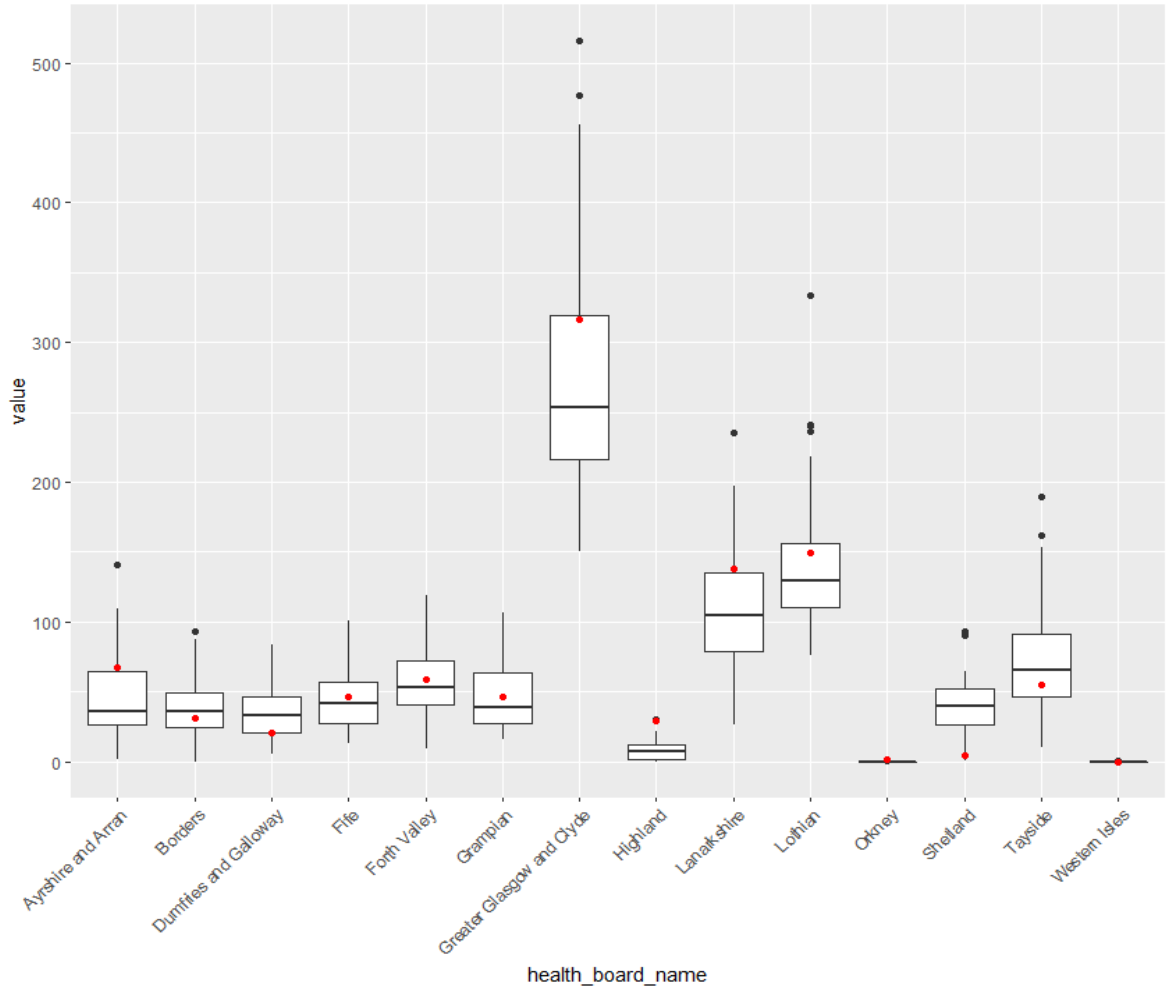

Figure 5. Distribution of simulation outcomes compared to the recorded number of COVID-19 related deaths as of 15th April 2020 (red dots).

The total COVID-19 related deaths are shown for all health boards in Figure 4. Because the model uses a single set of parameters to describe all of Scotland (i.e. it is assumed that the disease is transmitted in the same way across Scotland subject to variation in population density, age and differences in commuter patterns), it is expected that there is some variation in the way the model fits the data. However, comparison to a model fit without the health index shows a substantially improved fit when it is included (not shown). The best fit model shows a considerable influence of health index on mortality; the most likely impact is that the Health Board with the highest index will have, on average, a reduction of approximately 7% in mortality compared to the Health Board with the lowest health index.

2.iii impact of lockdown on COVID-19 spread.

Lockdown restrictions are likely to impact both the local spread of infection (via social distancing measures) and the spatial spread of infection (indirectly, social distancing means reductions in the number of long distance contact events). We illustrate the impact of lockdown on spatial spread, by considering the likely extent of disease should lockdown restrictions have been imposed two weeks prior to the actual date (23rd March 2020) which is approximately the time of the announcement of the first death due to COVID-19 in Scotland. Using our approach to simulating the effect of lockdown, as seen in Figure 5, the trajectory of simulated mortality remains consistent with the observed mortality rate until April 27th (the last date for the available data at point of fitting). In this version, if we assume that the reproduction rate declines below one from two weeks after the

imposition of lockdown (the same assumption as for the baseline scenario), we predict on average 581 deaths (95% of simulations within 377 to 1010 deaths) by 26th April 2020, compared to 2722 (range 1294 to 4050) in the baseline scenario (observed number is 2,795, assuming that all deaths occur in the week prior to the week the death is registered in).

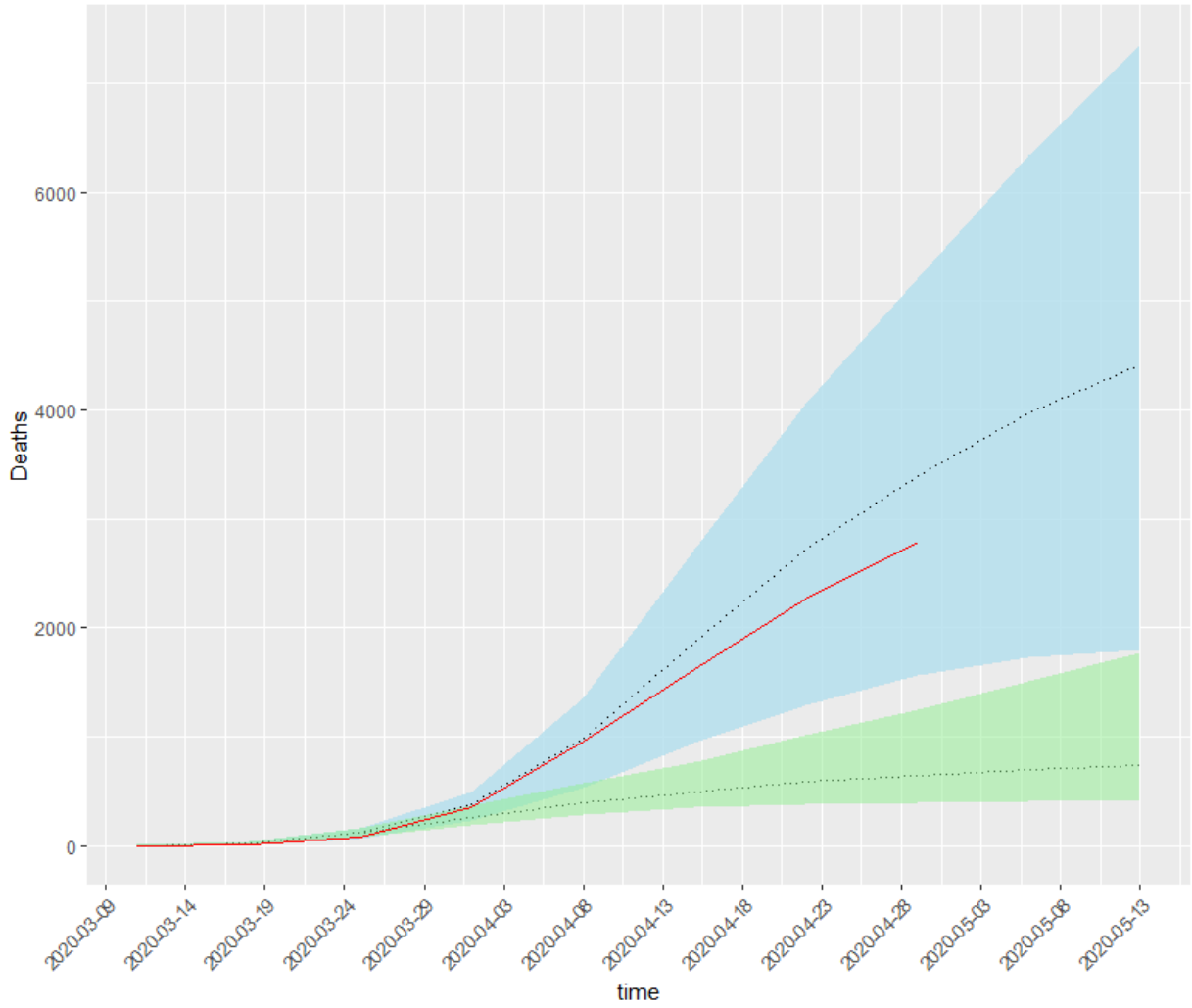

Figure 6. comparison of number of COVID-19 related deaths in early lockdown scenario (95% C.I. in green) to baseline (95% C.I. in blue).

The reduction in numbers also results in a reduction in geographical spread with many fewer OAs affected by COVID-19 mortality in the early lockdown scenario (figure 6). While these do not represent actual directions of spread (as the inference uses health board level statistics, these sub-health board level representations do not represent true distributions of cases) they serve to illustrate the role that movements between OAs play in the simulation.

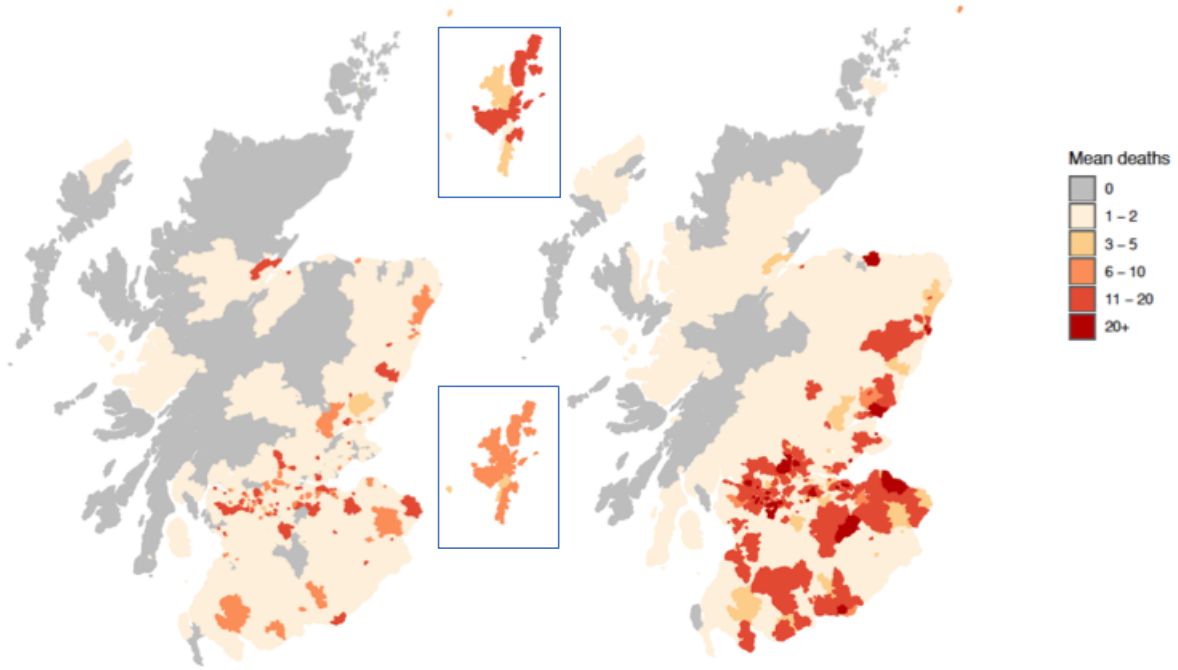

Figure 7. Mean number of deaths per OA, averaged over 50 simulations for early lockdown (L) and lockdown as it occurred (R). Deaths in Shetland shown in the inset with early lockdown at the top, and at bottom, as lockdown occurred.

3. Discussion.

In this report we to present the first description of the structure of SCoVMod, an individual-based model of COVID-19 spatio-temporal dynamics in Scotland. The model is shown to fit within acceptable limits over the course of the epidemic that is fitted, including when considering health board level statistics. To allow for the many simulations required for inference problems, the underlying population structure built into the current version of SCoVMod is fundamentally simple (though computationally intensive), with abstract work and home structures. The details of population structure is, however, implied in the model input and therefore is highly flexible to different scenarios. This allows, for example, for a much more detailed representation of home, work and recreation patterns, provided there are data to support this.

By embedding within the model the health index, taken from the SIMD averaged over all health boards, we obtain a direct estimate of the additional mortality related to COVID-19 in deprived areas, and correcting for possible correlations between deprivation and transmission dynamics. This illustration needs to be further investigated by a deeper interrogation of more finely grained datasets however even with the relatively coarse grained scale of health boards, still show that mortality in the least deprived health board is approximately 30% of that in the most deprived. Such substantial differences, built into future models, will provide a much more refined assessment of potential health burdens and risks associated with geographical spread.

By combining the impact of reduced work-related travel, with observations regarding reductions in the reproduction number of COVID-19. The two factors are of course not completely separable - reduction in long distance travel and social distancing measures go hand in hand. However our simple approach could be useful, for example, to strategically examine trade-offs between travel related restrictions, and social distancing when evaluating future releases from lockdown. The direct interpretation of the lockdown scenarios and in particular future projections must be viewed with some caution. The effectiveness of lockdown will vary in space and time, and due to reasons such as human behaviour (e.g. 'lockdown fatigue", or uncertainty over the exact control measures). On the

other hand, reduced numbers of cases will reduce spread, and therefore logistical burden, with possibly more resources available and the burden on care homes, hospitals and ICUs reduced. These potentially counterbalancing factors would of course have to be considered in more detail however our comparison can form the basis of evaluation of the relative benefit of timing of restrictions.

While this version of SCoVMod provides a useful environment to explore the implications of spatio-temporal spread of COVID-19 (and by extension, other highly infectious viral diseases with substantial mortality), there are several extensions now under consideration, including improved age dependent contact, finer resolution scale with defined workplaces, and improved scaling of movement patterns to consider not just changes in volume but also distance of movements.

Acknowledgements.

Samantha J. Lycett (both Roslin Institute, University of Edinburgh), for helpful advice on the model development, K. Atkins (Usher Institute, University of Edinburgh) for input on model development and advice on interpretation of outputs. Prof. C. Robertson (Strathclyde University, Health Protection Scotland) for advice on scenarios. Funding notes: this work has been funded by Roslin ISP2 (theme 3) - BBS/E/D/20002174, Wellcome Trust grant 209818/Z/17/Z, BBSRC grant BB/P010598/1

Appendix I. Model Equations.

For individuals in each OA, the force of infection $\Lambda(t)$ is given by:
Where:

$$\Lambda_i(t) = -\Bigg[(\beta_N + \beta_D)\sum_{a\in Y,E}\left(yI^M{}_{i,a} + I^S{}_{i,a}\right) + \beta_N\left(yI^M{}_{i,A} + I^S{}_{i,A}\right) + \beta_D I^S{}_{i,A} + \beta_D\left\{\left\{\sum_j\left(1-\sum_j x_{ij}\right)\left(yI^M{}_{\mathrm{i},A}\right)\right\} + \sum_j\left\{x_{ij}\left(\sum_k(1-x_{jk})\left(yI^M{}_{j,A}\right) + I^S{}_{j,A} + \sum_{a\in Y,E}\left(yI^M{}_{\mathrm{j},a} + I^S{}_{\mathrm{j},a}\right)\right)\right\}\right\}\Bigg]$$

With

$$\beta_N = nighttime\ tranmission\ rate$$
$$\beta_D = daytime\ tranmission\ rate$$
$$a = age\ class\ (Y\colon young, E\colon elderly)$$
$$y = reduction\ in\ transmission\ for\ mildly\ infected\ individuals$$
$$I^M_{i,a} = mildly\ infected\ in\ location\ i\ with\ age\ class\ 'a'$$
$$I^S_{i,a} = severely\ infected\ in\ location\ i\ with\ age\ class\ 'a'$$
$$x_{ij} = proportion\ of\ individuals\ commuting\ from\ location\ i\ to\ location\ j$$
$$a = age\ class\ (Y\colon young, E\colon elderly)$$
$$y = reduction\ in\ transmission\ for\ mildly\ infected\ individuals$$

In the compartmental model are infection classes *S* (susceptible), *E* (exposed), I^M (mildly infected), I^S (severely infected), and *H* (hospitalised). Model equations for individuals residing in one OA (labelled *i*) and for age class *A'* are therefore:

$$\frac{dS_{i,A'}}{dt} = -\Lambda_i(t)S_{i,A'}$$
$$\frac{dE_{i,A'}}{dt} = \Lambda_{i,A'}(t)S_{i,A'} - \gamma E_{i,A'}$$
$$\frac{dI^M{}_{i,A'}}{dt} = \gamma E_{i,A'} - (\theta + \rho_M)I^M{}_{i,A'}$$
$$\frac{dI^S{}_{i,A'}}{dt} = \theta I^M{}_{i,A'} - \left(\rho_S + \mu_{s,A'} + \eta\right)I^S{}_{i,A'}$$
$$\frac{dH_{i,A'}}{dt} = \eta I^S{}_{i,A'} - \left(\rho_H + \mu_{H,A'}\right)I^S{}_{i,A'}$$
$$\frac{dR_{i,A'}}{dt} = \rho_M I^M{}_{i,A'} + \rho_S I^S{}_{i,A'} + \rho_H \mathrm{H}_{i,A'}$$

Where the transition rates are given by:

- γ for *E* to I^M,
- ρ_M for I^M to *R*,

- ρ_S for I^S to R,
- ρ_H for H to R
- θ for I^M to I^S,
- ρ_S for I^S to R
- η for I^S to H,

and the age-dependent mortality rates are given by the values of μ.

NB: In order to improve the efficiency of the inference, movements of commuters between OAs were batched into groups of 5, with movements between OAs of fewer than five individuals per day, retained at a proportionate rate (i.e. 0.80 of all movements between OAs involving 4 individuals were retained, with the remaining discarded at random). While this reduces the overall network link density, the effect on transmission dynamics is expected to be small. We note that this means that interpretation of the combined β_N and β_D must be made with caution.

Appendix II. Generate of commuter movement patterns

From the current population estimates we draw the number of individuals whose primary residence is mapped onto OA[25], with their age group. The smallest geographic unit provided publicly by NRS is the Intermediate Zone (IZ)[26], of which there are approx. 1,200 units in Scotland each with a population of 2,500–6,000 household residents. We refine this, by synthetically distributing individuals down to Census OAs, of which there are approx. 46,000, each with a household population of 100–500. The total population of Scotland from this Census is 5438054 (Young, 919,580; Adult 3,492,421; Elderly 1,026,053). Of the adults, 1,960,712 commute to work (reduced to 647, 034 under lockdown (see details below).

The data for assignment of individuals to work locations is drawn from the NRS Census Flows data[27], Table WU01UK, which provides origin/destination workplace data for the population from the 2011 census. This is also provided at IZ level, which we distribute to OA level.[28]

The data for adjusting daily movements for the period after lockdown is taken from Google's Community Mobility Reports[29], which provide estimates of the proportionate decrease in mobility against a pre-lockdown baseline.

The population and census data were retrieved on 1st April 2020, and the mobility data is updated every few days (last update 1st May 2020).

1.iv Commuter patterns.

To model commuter patterns between OAs we construct a commuter network consistent with commuter patterns and density of working age adults recorded in the Scotland census.

1.iv.a Network patterns.

---

25 Geographic units defined by the ONS: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

26 NRS definition of geographic units: https://statistics.gov.scot/atlas/resource?uri=http://statistics.gov.scot/id/statistical-geography/S92000003

27 Office for National Statistics, 2011 Census: Special Workplace Statistics (United Kingdom) [computer file]. UK Data Service Census Support. Downloaded from: https://wicid.ukdataservice.ac.uk

28 https://www.scotlandscensus.gov.uk/variables-classification/sns-data-zone-2011

29 Google Community Mobility Reports: https://www.google.com/covid19/mobility/

Sources from the 2011 Scotland census[30] were combined to create a semi-synthetic network of work-related movements between households and workplaces. The census provides the number of individuals commuting between 'place of residence' and 'place of work' at a larger scale than the OA patches used in our model (Table ID WU01SC_IZ2011_Scotland). To generate commutes between the smaller 'OA' scale we assigned them a residence OA from within their 'place of residence' selected randomly with probability proportional to its working-age household population (Table ID LC1109SC), and a workplace OA from within their 'Place of work' selected randomly with probability proportional to the workforce population (Table ID WP101SCoa).

1.iv.b Movement simulation.

The purpose of the movement simulator ("moveSim") is to provide population and movement input to SCoVMod. Input is in the form of a national population, drawn from census data, and a set of movements—where each individual travels in each time period—drawn from census flows data and modified by mobility data. We assume in this model that the pattern of commuting captures the most important features of . Also, after lockdown, this effect will be enhanced; with evidence from other countries of non-work activity occurring at both reduced volumes [REF google] and at a reduced spatial distance.[31]

Population

The first output of moveSim is a national population of individuals, each with a unique ID, a starting location, and an age group. At this point we also decide if each individual goes to work and where this occurs, assign them a work location.

Algorithm

The algorithm for assignment of individuals to home locations and age categories applies a unique ID for each individual in the national population. Home locations and age groups are assigned proportionately to the estimated OA population size.

An individual's workplace is assigned by distributing a proportion of the population of each location to each work location, weighted by the proportion of individuals from each home location in the census flows data who work in another location. For each origin o and destination d we assign a weight $w_{o,d}$ from the census flow data: $w_{o,d} = \frac{n_{o,d}}{t_o}$ where $n_{o,d}$ is the total number of people who move from o to d to work, and t_o is the total number who move from origin o to any location for work.

We take the individuals of each home location if they are eligible to work (total n_o); in this case we assume all individuals of adult age 16–65. Each destination is assigned to $n_o \times w_{o,d}$ of these individuals. The individuals who remain have no assigned workplace—either they do not work, or they work within their home location.

Output format

The population generator of moveSim outputs two tables: the first with *PersonID, Origin, Age* as input to SCoVMod, and the second *PersonID, Origin, Age, Worker, Destination* as input to the moveSim movement simulator. *PersonID* is a unique integer identifier, *Origin* and *Destination* are string location identifiers, *Age* is a categorical identifier in {*Young, Adult, Elderly*}, and *Worker* a boolean indicating an eligible worker.

[30] Census data: https://www.scotlandscensus.gov.uk/ods-web/data-warehouse.htmla

[31] https://www.covid-19-mobility.org

Movement

The movement simulator produces input for SCoVMod, in the form of the set of individuals who move from each location in each time step of the simulation. In this case, we use two time steps per day. In the first, workers move to work, and in the second they return home. The simulation is generated stochastically, with a Poisson distributed number of workers moving from each origin to each destination per day, distributed according to the census flows and weighted by population as described above. The volume of movement is reduced uniformly across the population according to the proportional decrease provided by the mobility data.

We also introduce an optimisation to reduce the number of movements that need to be handled by SCoVMod. The number of movements are trimmed to one in five and therefore daytime transmission rates are correspondingly assumed to be per 5 infected individuals. In simulation, there will be many OAs where fewer than five infected individuals would move to them in a given time step, and thus this process of movement thinning would result in many locations not being exposed to infection that would have been with the full movement pattern. Given the low level of infection over the considered scenarios, we assume that the trade-off between increased transmission rate per movement, and reduced movements, will have negligible impact on outcomes.

Algorithm

For each day of the simulation we consider two time steps: a *day* step where individuals can move to their place of work, and a *night* step where those individuals move back to their home location.

In each day step, we take each destination location d. Let λ_d be the number of eligible workers who may move to the destination location. For each day the sampled number who move s is drawn from a Poisson distribution:

$$s \sim Poisson(\lambda_d)$$

The sampled number of moves s is then scaled according to the per cent change in mobility m for the given day:

$$s_m = \lfloor s(1 + (m/100)) \rfloor$$

The number of moves is then trimmed 4 in 5 by drawing from a Binomial distribution:

$$s_{mt} \sim B(s_m, \frac{1}{5})$$

If the sampled number of workers s_{mt} is less than or equal to the number of workers who may normally move to destination d, then those who move are sampled randomly from those who may normally move. However, if s_{mt} is greater than the number of workers who may normally move to d, then the additional workers are drawn randomly from workers who have no assigned destination location. The sampled PersonIDs for each destination location are then collected for output, for each day. For each night of the simulation, the workers who moved in the day step are moved back to their origin location.

Output format

Each time step of the simulation is output in JSON format for input to SCoVMod. Each time step contains the set of destination location IDs, each containing the set of PersonIDs who move to them.

Appendix III. Evaluation of validity of google mobility data in rural areas.

Reduction in mobility is estimated based on data provided by Google.[32] In the documentation, it is stated that "Location accuracy and the understanding of categorized places varies from region to

[32] https://www.google.com/covid19/mobility/index.html?hl=en

region, so we don't recommend using this data to compare changes between countries, or between regions with different characteristics (e.g. rural versus urban areas)."[33]

In order to provide confidence that inclusion of reductions across regions is appropriate, we here assume that urban areas such as Glasgow and Edinburgh are likely to be well represented, but that rural areas may be less so. To check this, we compare an independent dataset on independent sailings and passenger numbers for ferry services run by Caledonian MacBrayne, who operate all ferry services in the west of Scotland. A comparison of data from 2019 to 2020 and to Google Mobility data, shows a strong fidelity between the two datasets, as well as a substantial reduction in activity at point of lockdown. The similarity prior to lockdown between 2019 and 2020 also suggests that patterns of increased summer activity are unlikely to have had strong influences on our assumptions regarding commuter movements, at least in this area.

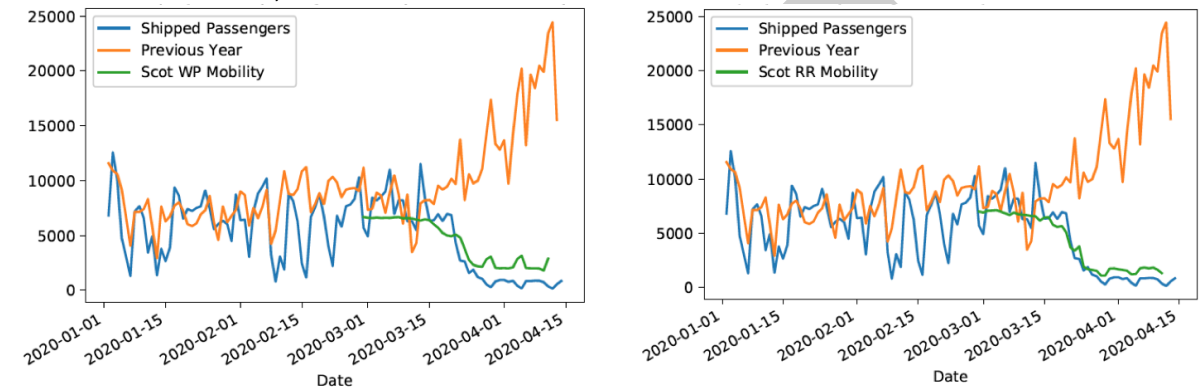

Figure III.i. Comparison of Google mobility data for Scotland to CalMac Ferry records. (L) Comparison to workplace mobility (R) Comparison to Recreation mobility. The comparison is relative to the mean value prior to lockdown on March 23rd, 2020.

Appendix IV. References for parameter estimates and priors (cf. Table 1).

1. Arenas, A. et al. A mathematical model for the spatiotemporal epidemic spreading of COVID19. medRxiv 2020.03.21.20040022 (2020). doi:10.1101/2020.03.21.20040022
2. Davies, N. G. et al. The effect of non-pharmaceutical interventions on COVID-19 cases, deaths and demand for hospital services in the UK: a modelling study. medRxiv 2020.04.01.20049908 (2020). doi:10.1101/2020.04.01.20049908
3. Ferguson, N. M. et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial.Ac.Uk 3–20 (2020). doi:10.25561/77482
4. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med. (2020). doi:10.1038/s41591-020-0869-5
5. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Science 3221, 1–9 (2020).
6. Qifang, B. et al. Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. medRxiv 1–22 (2020).
7. Zhang, J. et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect. Dis. 3099, 1–10 (2020).
8. Sanche, S. et al. The Novel Coronavirus, 2019-nCoV, is Highly Contagious and More Infectious Than Initially Estimated. (2020).
9. Famulare, M. nCoV: incubation period distribution. (2020). Available at: https://institutefordiseasemodeling.github.io/nCoV-

[33] https://www.google.com/covid19/mobility/data_documentation.html?hl=en

public/analyses/individual_dynamics_estimates/nCoV_incubation_period.html. (Accessed: 7th May 2020)
10. Linton, N. M. et al. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. J. Clin. Med. 9, 538 (2020).
11. Lauer, S. A. et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann. Intern. Med. 2019, (2020).
12. Sanche, S. et al. High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2. Emerg. Infect. Dis. 26, (2020).
13. Tindale, L. et al. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. medRxiv 2020.03.03.20029983 (2020). doi:10.1101/2020.03.03.20029983
14. Jing, Q. et al. Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study. medRxiv 2020.03.06.20032417 (2020). doi:10.1101/2020.03.06.20032417
15. Wang, D. et al. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA - J. Am. Med. Assoc. 323, 1061–1069 (2020).
16. Deng, X. Case fatality risk of novel coronavirus diseases 2019 in China. medRxiv 1–13 (2020). doi:10.1101/2020.03.04.20031005
17. Liu, T. et al. Time-varying transmission dynamics of Novel Coronavirus Pneumonia in China. bioRxiv 2020.01.25.919787 (2020). doi:10.1101/2020.01.25.919787
18. Arentz, M. et al. Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State. JAMA 323, 1612–1614 (2020).
19. Han, Y. nan et al. A comparative-descriptive analysis of clinical characteristics in 2019-coronavirus-infected children and adults. J. Med. Virol. (2020). doi:10.1002/jmv.25835
20. Pung, R. et al. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. Lancet 395, 1039–1046 (2020).
21. Dorigatti, I. et al. Report 4 : Severity of 2019-novel coronavirus (nCoV). Imp. Coll. London COVID-19 Response Team 1–12 (2020). doi:10.25561/77154

Appendix VI. BBC Disclosure, 11th May 2020. Effect of early lockdown.

An earlier version of this model which did not include consideration of health index, was used in the BBC One "Disclosure" programme aired at 8:30 pm on Monday, 11th May 2020. While the model results from the refined version are slightly different, the overall conclusions, and estimated number of deaths averted are highly consistent. Using the earlier fitted model, we also explore the impact of an early lockdown, starting at the beginning of the simulation period on March 8th, two weeks prior to the actual lockdown rate. In this version, if we assume that the reproduction rate declines below one from two weeks after the imposition of lockdown (the same assumption as for the baseline scenario), we predict on average 577 deaths (95% of simulations within 343 to 1141 deaths) by 26th April 2020, compared to 3259 (range 1736 to 6983) in the baseline scenario (observed number is 2,795, assuming that all deaths occur in the week prior to the week the death is registered in).

Appendix VII. Author contributions.

Banks C.J.*, Colman E. *, Doherty T. *, Tearne O. *, Balaz, D., Arnold M., Beaunée, G., Bessell P., Enright J., Rossi, G., Ruget A.-S., Kao R.R.†

CB and EC generated synthetic movement patterns, TD and OT developed and unit tested the code, and developed the ABC-SMC inference scheme. CB provided the Calmac Ferry data comparison. MA provided oversight at APHA. GB advised on ABC-SMC inference, PB conducted the statistical analysis of SIMD data, JE contributed to generation of synthetic movement patterns and provided input on model structure throughout the project. DB provided code review. GR proposed parameter point estimates and prior distributions, ASR generated plots and conducted statistical analyses, RRK conceived the project, and supervised the overall project. CB, EC, TD, and RRK wrote the initial manuscript draft. All authors have read and critiqued the manuscript. RRK is the author for correspondence.