COVID-19 in Scottish care homes: A metapopulation model of spread among residents and staff

Matthew Baister [1◕], Ewan McTaggart[1◕], Paul McMenemy,[2], Itamar Megiddo[3], Adam Kleczkowski[1],

1 Department of Mathematics & Statistics, University of Strathclyde, Glasgow, UK
2 Department of Computing Science and Mathematics, University of Stirling, Stirling, UK
3 Department of Management Science, University of Strathclyde, Glasgow, UK

◕These authors contributed equally to this work.

Abstract

Care homes in the UK were disproportionately affected by the first wave of the COVID-19 pandemic, accounting for almost half of COVID-19 deaths over the course of the period from 6th March – 15th June 2020. Understanding how infectious diseases establish themselves throughout vulnerable communities is crucial for minimising deaths and lowering the total stress on the National Health Service (NHS Scotland). We model the spread of COVID-19 in the health-board of NHS Lothian, Scotland over the course of the first wave of the pandemic with a compartmental Susceptible - Exposed - Infected reported - Infected unreported - Recovered - Dead (**SEIARD**), metapopulation model. Care home residents, care home workers and the rest of the population are modelled as subpopulations, interacting on a network describing their mixing habits. We explicitly model the outbreak's reproduction rate and care home visitation level over time for each subpopulation, and execute a data fit and sensitivity analysis, focusing on parameters responsible for intra-subpopulation mixing: staff sharing, staff shift patterns and visitation. The results suggest that hospital discharges were not predominantly responsible for the early outbreak in care homes, and that only a few such cases led to infection seeding in care homes by the 6th of March Sensitivity analysis show the main mode of entry into care homes are infections by staff interacting with the general population. Visitation (before cancellation) and staff sharing were less significant in affecting outbreak size. Our model suggests that focusing on the protection and monitoring of staff, followed by reductions in staff sharing and quick cancellations of visitation can significantly reduce future infection attack rates of COVID-19 in care homes.

Introduction

The outbreak of the SARS-CoV-2 induced disease (COVID-19) pandemic has had a profound impact, causing 3.7 million deaths by early June 2021 and economic hardship globally [1]. The care home population suffered a disproportionate amount of COVID-19 related deaths. From the week ending 13th March 2020 to the week ending 26th June 2020 (the "first wave"), 54,510 deaths were associated with COVID-19 in the UK, 40% of which were among care home residents [2]. The COVID-19 pandemic has highlighted the vulnerability of care homes, as their resident population is elderly and often suffers from several co-morbidities [3], their systems have not been developed with infection prevention and control (IPC) in mind, and their IPC guidelines have been borrowed from hospitals - a completely different setting [4]. Networks of care homes are an ecosystem connected by staff working across facilities. These connections increase the risk of COVID-19 ingress into care homes, and to protect their vulnerable community, we need to understand the ecosystem dynamics.

Care homes and their residents are enclosed societies, isolated to some extent from the general population. Their connection to broader society primarily consists of interaction with staff and visits from the general population. Care home staff potentially play a vital role in COVID-19 introduction and spread throughout the care home population. Firstly, staff exposure to infection from the general population can establish an outbreak in a home. Secondly, some staff work across multiple homes - a concept we refer to as staff sharing. Staff acting as a bridge between care homes and the general population and staff sharing induces a network, connecting care homes in a given community via their workers. This creates the potential for COVID-19 to spread from one home to another; hence, investigation of this pathway is important. We find it natural to describe this using a heterogeneous patch size metapopulation model framework.

Very few models explore COVID-19 transmission at a community level and explicitly include the unique dynamics in care homes. For example, in [5, 6] agent-based models (ABMs) of single homes are used to investigate the impact of testing strategies on the disease burden. A report by Nyguen et al. [7] uses an ABM to investigate the impact on care home residents of various vaccine coverage, and reducing the weekly testing of staff. However, the models in [6, 7] do have an external force of infection (FOI) from the community, based on prevalence data, representing staff interaction with the community and visitors. These models [5–7] do not assess the relative impact of the different COVID-19 pathways into care homes. Nguyen et al. [8] extend [6, 7], using a hybrid ABM-System Dynamics model, to explore the conditions under which visitation, heterogeneous care homes sizes, and the cohorting of residents impacts COVID-19 outbreak severity.

Rosello et al. [9] model an individual care home with a stochastic compartmental model, using multiple FOI's to capture COVID-19 pathways, including visitors, hospital discharges, staff working at other homes, and staff infections from the community. They find that importations of infections by staff from the community are the main driver of outbreaks, and importation by visitors or from hospitals is rare, but do not explicitly model disease spread throughout a network of care homes. In [10] individual care homes and the general public are independent, deterministic SEIR models, with a stochastic external FOI connecting the general public to each home. This FOI depends on the prevalence of COVID-19 in the general public, and the size of each home. Transmission rates in homes and in the general public do not vary over

time. In [11], two weakly-coupled SEIR sub-models with time-dependent transmission rates define the dynamics; one sub-model describes the general public and one describes all care home residents in Stockholm as a single homogeneous group. Again, a single FOI acts on the residents to capture infections from staff and visitors. The models [10, 11] do not differentiate between, and therefore allow comparison of, the COVID-19 pathways into care homes. Bunnik et al. [12] use a compartmental metapopulation model to explore the trade-offs between increasing protection for a "vulnerable" population and relaxing restrictions for the "non-vulnerable" after the first lockdown in Scotland. They use time-dependent transmission rates with three metapopulation groups; vulnerable, shielders and general public. We extend and apply the methodology of [12] in our model, investigating protection to a vulnerable group (care home residents) in ways other than shielding.

We construct a **SEIARD** compartmental metapopulation model to describe the first wave of COVID-19 in a health board in Scotland. The population is divided into groups of care home residents, staff, and general public. Our care home resident group are not a single homogeneous unit as in [11, 12] but are separate units, creating a refined spatial/geographic structure. These units are not independent as in [10] but are linked by a staff sharing network which, to our knowledge, is unique. We calibrate this model to data from the NHS Lothian Health Board and explore the sensitivity of the results to changes in key parameters. We investigate the importation of infections by staff from the community, visitation, staff sharing, and additionally, we shed light on the exposure of care homes at the beginning of the first wave, e.g., via hospital discharges [13]. The aim is to identify and rank the modes of COVID-19 ingress into and throughout the susceptible care home community in order to improve future pandemic responses. Our model allows for this investigation by coupling the general public and individual care homes with the explicit movement of staff and visitors between the two.

Materials and methods

Mathematical model

We develop a deterministic **SEIARD** compartmental metapopulation model with heterogeneous subpopulation sizes. Each subpopulation consists of a host human population, categorised further into six compartments of COVID-19 infection status: Susceptible (**S**), i.e., everyone who is not infected; Exposed (**E**), those exposed to the virus (and infected) but not yet infectious; Infectious and reported (**I**), infectious individuals that have been identified with a positive test; Unreported infectious (**A**), infectious individuals that have not been identified with a positive test; Recovered (**R**), those who had COVID-19 and recovered; and Dead (**D**), those who died from their illness. Symptomatic and asymptomatic individuals are not modelled explicitly; instead, asymptomatic infections contribute towards a reduction in the reporting rates. This model is illustrated in Fig 1 (a).

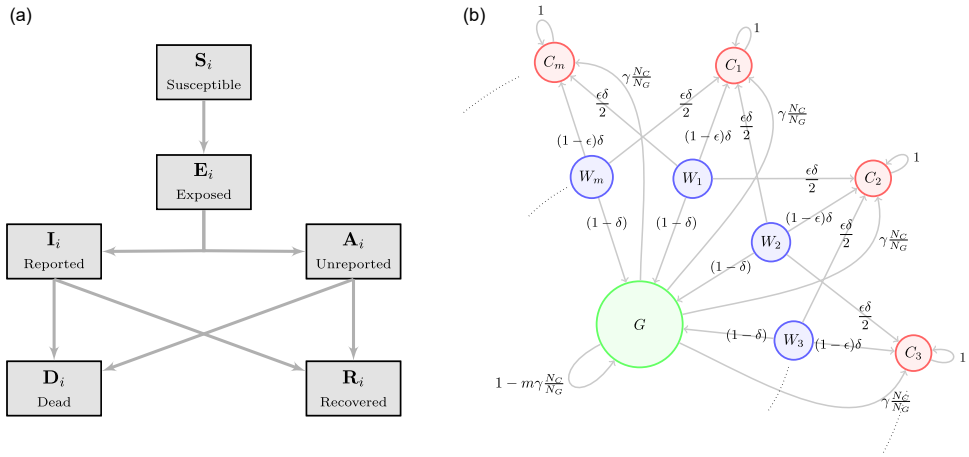

Fig 1. Schematics for the compartmental and metapopulation structure. (a): **SEIARD** compartmental structure of the model; **(b)**: Time-share network of interaction amongst subpopulations. Directed edge weights are t_{ik}, the proportion of people from subpopulation i who travel to mix at effective population k.

The metapopulation structure represents the population of the NHS Lothian health board in Scotland. We distinguish between care home residents, care home workers and the general population, modelling the $m = 109$ care homes for older adults in NHS Lothian [14]. The jth home has a resident subpopulation, C_j, with a corresponding care home worker subpopulation, W_j. The general population is encapsulated by the subpopulation G. Each care home includes the same number of residents, a simplifying assumption made due to lack of publicly available data on care home sizes in Lothian. We also assume the worker subpopulations are the same size as residents' [15].

Each node of the network, $i \in X := \{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$ with $|X| = n$, is described in terms of the **SEIARD** compartmental model with equations:

$$
\begin{aligned}
\frac{d\mathbf{S}_i}{dt} &= -\mathbf{S}_i \Lambda_i \\
\frac{d\mathbf{E}_i}{dt} &= \mathbf{S}_i \Lambda_i - \frac{\mathbf{E}_i}{\lambda} \\
\frac{d\mathbf{I}_i}{dt} &= r_i \frac{\mathbf{E}_i}{\lambda} - \frac{\mathbf{I}_i}{\tau} \\
\frac{d\mathbf{A}_i}{dt} &= (1 - r_i) \frac{\mathbf{E}_i}{\lambda} - \frac{\mathbf{A}_i}{\tau} \\
\frac{d\mathbf{R}_i}{dt} &= (1 - \mu_i) \frac{(\mathbf{I}_i + \mathbf{A}_i)}{\tau} \\
\frac{d\mathbf{D}_i}{dt} &= \mu_i \frac{(\mathbf{I}_i + \mathbf{A}_i)}{\tau}
\end{aligned}
\tag{1}
$$

Susceptibles in subpopulation i ($\mathbf{S}_i$), are infected with a FOI Λ_i, and moved to the exposed class ($\mathbf{E}_i$). After a non-infectious latent period of λ days, they become infectious, testing positive at a reporting rate of r_i. These identified infections move to the class $\mathbf{I}_i$. Hence, any unidentified infections, $\mathbf{A}_i$, occur at rate 1 - r_i. After τ infectious days, a person either recovers or dies at the rate μ_i. For simplicity, and considering the short span of time the model is designed to describe, non-COVID

related deaths are not considered. For similar reasons, we do not include a birth rate or admission of new residents to care homes from the general population.

We assume a constant reporting rate for care home residents ($r_{i\in\{C_1,\ldots,C_m\}} = r_C$), workers ($r_{i\in\{W_1,\ldots,W_m\}} = r_W$), and the general public (r_G). The parameters τ and λ describe the infectious period and latency period, respectively, and are assumed to be the same across all subpopulations. Mortality rates, μ_i, vary by subpopulation, reflecting the positive association of serious outcomes of COVID-19 with age [16]. As we are modelling over a period of 4 months (approx. first wave), and immunity after COVID-19 infection lasts as long as 5 months [17] [18], we do not consider a transition from Recovered to Susceptible.

We model visitation to each care home by multiplying the proportion of the population, N_C/N_G, that visit the care home, and their length of stay, $\gamma(t)$. The proportion remains constant over time while $\gamma(t)$ varies over time. Up until 13th March 2020, each resident has one visitor per day [6]. Then $\gamma(t)$ drops to 0, reflecting the policy change to essential visitation only [14, 19]. $\gamma(t)$ is described by the function logi(Ω_x), with $\Omega_x = \{t, \omega^x_{end}, \omega^x_{rate}, \omega^x_{low}, \omega^x_{high}\}$, defined below:

$$logi(\Omega_x) = \frac{(\omega^x_{high} - \omega^x_{low})}{(1+\exp(\omega^x_{rate}(t-\omega^x_{end})))(1+\exp(-\omega^x_{rate}(t+82)))} + \omega^x_{low}, \qquad (2)$$

with the shape of a sigmoidal logistic function. The value of 82 is used in the function so that when $t = 0$, $logi(\Omega_x) = \omega^{high}$. The function drops from ω^{high} to ω^{low} at a time controlled by ω^x_{end}, such that when $t = \omega^x_{end}$, $logi(\Omega_x) = (\omega^x_{high} + \omega^x_{low})/2$. The ω_{rate} parameter changes the gradient of the descent at $t = \omega_{end}$.

Thus, visitation rate is described by $\frac{N_C(0)}{N_G(0)}\gamma(t) = \frac{N_C(0)}{N_G(0)} logi(\Omega_\gamma)$. Given that visitation drops to 0 in the first 2 weeks of the simulation, the changes in population size over that time is negligible, hence we can keep the proportion of the population constant and control visitation by solely changing $\gamma(t)$.

A constant proportion of workers, δ, spend their time at care homes. With $\delta = 0.5$, half of a worker compartment, W_i, are at care homes, C_i, over the course of a day. The proportion of workers not at care homes, $1-\delta$, reside in the general population, G. During the first wave of COVID-19 in Scotland, there was both intra-organizational staff sharing between homes (i.e., staff who work at multiple homes belonging to the same care provider), as well as inter-organizational staff sharing (use of bank or agency staff) [20, 21]. Therefore, a constant proportion of each homes' assigned workers, ε, are exchanged between homes. We refer to this as *staff sharing*. We have made the simplifying assumption that the staff sharing network has a topology of a circle, whereby the shared staff for home j are split evenly between homes $j-1$ and $j+1$. We assume care home residents do not leave their homes.

Interaction across subpopulations is heterogeneous and is described in terms of time-sharing, determining proportions of subpopulations mixing in groups with each other. In the ith subpopulation there are $N_i(t) = \mathbf{S}_i(t) + \mathbf{E}_i(t) + \mathbf{I}_i(t) + \mathbf{A}_i(t) + \mathbf{R}_i(t)$ active individuals who can mix with others. The proportion from subpopulation i who travel to, and mix with, subpopulation k is t_{ik}. The effective population size of subpopulation k, given that others have travelled to it and some people from k have left, is $\hat{N}_k(t) = \sum_{j\in X} t_{jk} N_j(t)$. We assume these effective populations $\hat{N}_k(t)$ are well

mixed, so people who travel to each population can meet all others there. There are two types of effective populations; the care homes and the general population. Care home j, comprises $\hat{N}_{C_j}(t)$ people: its residents, its working staff, staff from other care homes, and visitors. The general population consists of $\hat{N}_G(t)$ people; this includes all the staff not at work and the non-visiting general population.

Our specific time-share assumptions are represented visually as a directed, weighted network in Fig 1 (b). The corresponding weighted adjacency matrix, the travel/time-share matrix, is $\boldsymbol{T} \in \mathbb{R}^{n\times n}$, whose $[i,j]^{th}$ element is t_{ij}. The rows and columns of $\boldsymbol{T}$ are in the order of $\{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$. $\boldsymbol{T}$ consists of the partitions $\{T_{CC}, T_{CW}, T_{WC}, T_{WW}, T_{CG}, T_{WG}, T_{GC}, T_{GW}\}$. To clarify notation: matrix $\boldsymbol{I}_m$ indicates the identity matrix of dimension m, matrix $[a]_{m\times m}$ indicates a matrix of dimension $m \times m$ with all entries a. Hence $\boldsymbol{T}$ and the subsequent sub-matrices are as follows:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T_{CC}} & \boldsymbol{T_{CW}} & \boldsymbol{T_{CG}} \\ \boldsymbol{T_{WC}} & \boldsymbol{T_{WW}} & \boldsymbol{T_{WG}} \\ \boldsymbol{T_{GC}} & \boldsymbol{T_{GW}} & \boldsymbol{T_{GG}} \end{bmatrix}_{n\times n}, \tag{3}$$

$$\boldsymbol{T_{CC}} = \boldsymbol{I}_m, \quad \boldsymbol{T_{CW}} = \boldsymbol{T_{WW}} = \begin{bmatrix} 0 \end{bmatrix}_{m\times m}, \quad \boldsymbol{T_{CG}} = \begin{bmatrix} 0 \end{bmatrix}_{m\times 1}, \quad \boldsymbol{T_{GW}} = \begin{bmatrix} 0 \end{bmatrix}_{1\times m},$$

$$\boldsymbol{T_{WG}} = \begin{bmatrix} 1-\delta \end{bmatrix}_{m\times 1}, \quad \boldsymbol{T_{GC}} = \begin{bmatrix} \frac{N_C(0)}{N_G(0)}\gamma(t) \end{bmatrix}_{1\times m}, \quad \boldsymbol{T_{GG}} = \begin{bmatrix} 1 - m\frac{N_C(0)}{N_G(0)}\gamma(t) \end{bmatrix}_{1\times 1},$$

$$\boldsymbol{T_{WC}} = \begin{bmatrix} (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} & 0 & \dots & 0 & \frac{\varepsilon\delta}{2} \\ \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} & \dots & 0 & 0 \\ 0 & \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & (1-\varepsilon)\delta & \frac{\varepsilon\delta}{2} \\ \frac{\varepsilon\delta}{2} & 0 & 0 & \dots & \frac{\varepsilon\delta}{2} & (1-\varepsilon)\delta \end{bmatrix}_{m\times m}. \tag{4}$$

Disease transmission in the model is assumed to be frequency-dependent. The FOI integrates which infections occur to whom, from whom and where the infection takes place, as in [22, 23]. The FOI acting on subpopulation i, Λ_i (see Equation 5), accounts for the different groups' people from i mix in over a day and who they encounter. It is most easily understood by considering $\Lambda_i \mathbf{S}_i$:

$$\Lambda_i \mathbf{S}_i = \sum_{k\in L_i} \frac{t_{ik}\mathbf{S}_i}{\hat{N}_k(t)} \sum_{j\in X} \beta_{ji}(t) t_{jk}(\mathbf{I}_j + \mathbf{A}_j) \tag{5}$$

The set of effective populations that subpopulation i travels to is L_i, consistent with the non-zero elements in the i_{th} row of the travel matrix $\boldsymbol{T}$. At effective population k, there is $t_{ik}\mathbf{S}_i$ susceptible individuals from i. At k there will also be $t_{jk}(\mathbf{I}_j + \mathbf{A}_j)$ infectious people from j who have travelled to k. The transmission rate between subpopulation j and i is $\beta_{ji}(t)$. Therefore,

$$\frac{t_{ik}\mathbf{S}_i}{\hat{N}_k} \beta_{ji}(t) t_{jk}(\mathbf{I}_j + \mathbf{A}_j)$$

is the number of new daily infections in i caused by people from j at the effective population k.

The transmission rates $\beta_{ji}(t)$ allow us to represent heterogeneous interaction patterns of individuals between and within different subpopulations. They incorporate the transmission dynamics of COVID-19 changing over time and location, for example, through lockdowns or other changes in behaviour [22]. We write $\beta_{ji}(t) = \frac{R(t)_{ji}}{\tau}$, describing the transmission rate $\beta_{ji}(t)$ between subpopulations j and i, with the reproduction rate, $R(t)_{ji}$, divided by the infectious period, τ. The contact rate and infection probability between subpopulations i and j is captured by $R(t)_{ji}$. We assume only the transmission rates between and within the subpopulation types (residents C, workers W, general public G) differ. Therefore, the transmission rates are arranged in a symmetric partitioned matrix $\boldsymbol{\beta} \in \mathbb{R}^{n\times n}$ whose $[j,i]^{th}$ element is $\beta_{ji}(t)$. The rows and columns of $\boldsymbol{\beta}$ are in the order of $\{C_1, C_2, ..., C_m, W_1, W_2, ..., W_m, G\}$. $\boldsymbol{\beta}$ contains block sub-matrices;

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta_{CC}} & \boldsymbol{\beta_{WC}^T} & \boldsymbol{\beta_{GC}^T} \\ \boldsymbol{\beta_{WC}} & \boldsymbol{\beta_{WW}} & \boldsymbol{R_{GW}^T} \\ \boldsymbol{\beta_{GC}} & \boldsymbol{\beta_{GW}} & \boldsymbol{\beta_{GG}} \end{bmatrix}$$

$$\boldsymbol{\beta_{CC}} = \Big[\beta_C(t)\Big]_{m\times m},\ \boldsymbol{\beta_{WC}} = \Big[\beta_C(t)\Big]_{m\times m},\ \boldsymbol{\beta_{WW}} = \Big[\beta_W(t)\Big]_{m\times m}$$

$$\boldsymbol{\beta_{GC}} = \Big[\beta_C(t)\Big]_{1\times m},\ \boldsymbol{\beta_{GW}} = \Big[\beta_G(t)\Big]_{1\times m},\ \boldsymbol{\beta_{GG}} = \Big[\beta_G(t)\Big]_{1\times 1}$$

The matrix notation above is the same as for the travel matrix $\boldsymbol{T}$. For simplicity, we have assumed that the resident-resident, worker-resident, and general population-resident transmission rates are equal. Similarly, we assume the general population-worker and general population-general population transmission rates are the same. The transmission rates are described by:

$$\beta_C(t) = \frac{logi(t, \omega^c_{end}, \omega^c_{rate}, \omega^c_{low}, \omega^c_{high})}{\tau},$$
$$\beta_W(t) = \frac{logi(t, \omega^c_{end}, \omega^c_{rate}, \omega^c_{low}, (\omega^c_{high} + \omega^G_{high})/2)}{\tau}$$
$$\beta_G(t) = \frac{logi(t, \omega^G_{end}, \omega^G_{rate}, \omega^G_{low}, \omega^G_{low}, \omega^G_{high})}{\tau}$$

Where the logi function, Equation 2, models the reproduction rate. To simplify and to reduce the number of parameters, we relate the reproduction rate for workers in terms of the residents and general population. As care home workers balance their time between care homes and the general population, we assume the workers pre-lockown maximum reproduction rate is the average of the care homes and general populations, $\omega^W_{high} = (\omega^C_{high} + \omega^G_{high})/2$. We assume the reproduction rate for workers and residents drops at the same time, and to the same value.

Model calibration process

We used data from the network of care homes in NHS Lothian [14] complemented by Public Health Scotland Open Data, breaking down COVID-19 cases and deaths per health board [24, 25], to inform and calibrate our model. Parameters were found using

a mixture of methods, as indicated in Table 1, including literature search, sensitivity of results, and rigorous fit based on minimising the sum of squares of residuals.

Table 1. Parameter definitions, alongside their base case values and source.

Parameter	Description	Value	Source
ε	Staff sharing	0.4	Data fit
δ	Proportion of workers at care home	0.5	Assumption[a]
$r_{i\in\{C_1,\ldots,C_m\}} = r_C$	Reporting rate for residents	0.53	Estimated [14, 28–30]
$r_{i\in\{W_1,\ldots,W_m\}} = r_W$	Reporting rate for workers	0.52	Estimated [14, 24, 27, 30, 31]
r_G	Reporting rate for general public	0.077	Estimated [24, 27, 31]
$\mu_{i\in\{C_1,\ldots,C_m\}} = \mu_C$	Death rate for residents	0.25	Estimated [14, 28–30]
$\mu_{i\in\{G,W_1,\ldots,W_m\}} = \mu_G$	Death rate for general public (and workers)	0.017	Estimated [24, 25, 27, 31]
τ	Infectious period	7 days	[32]
λ	Latent period	5.8 days	[33]
m	Number of care homes	109	[14]
$N(0) = \sum_i N_i(0)$	Total initial population	907,580	[27]
$N_{C_i}(0)$	Initial resident subpopulation size	48	[14]
$N_{W_i}(0)$	Initial worker subpopulation size	48	[15]
$N_G(0)$	Initial general public subpopulation size	897,116	Estimated [14, 15, 27]
$\omega^{\mathbf{C}}_{\mathbf{end}}$	Timing of R_t descent for residents and workers	42 days	Manually fitted [b]
$\omega^{\mathbf{C}}_{\mathbf{rate}}$	Rate of descent of R_t for residents and workers	0.5	Assumption[c]
$\omega^{\mathbf{C}}_{\mathbf{low}}$	Post-descent R_t for residents and workers	0.6	Assumption[d]
$\omega^{\mathbf{C}}_{\mathbf{high}}$	Pre-descent R_t for residents	4.7	Data fit
$\omega^{\mathbf{G}}_{\mathbf{end}}$	Timing of R_t descent for general population	22 days	Manually fitted [b]
$\omega^{\mathbf{G}}_{\mathbf{rate}}$	Rate of descent of R_t for general population	0.5	Assumption[c]
$\omega^{\mathbf{G}}_{\mathbf{low}}$	Post-descent R_t for general population	0.6	[34]
$\omega^{\mathbf{G}}_{\mathbf{high}}$	Pre-descent R_t for the general population	4.1	Data fit
$\omega^{\gamma}_{\mathbf{end}}$	Timing of descent for visitation	10 days	[14, 19]
$\omega^{\gamma}_{\mathbf{rate}}$	Rate of descent of visitation	3	[14, 19]
$\omega^{\gamma}_{\mathbf{low}}$	Post-descent value for visitation	0	Assumption[e]
$\omega^{\gamma}_{\mathbf{high}}$	Pre-descent R_t for visitation	0.083	Data fit
H_{seeded}	Number of homes seeded	4	Data fit
$E_G(0)$	Initial general population infections	120	Data fit

[a] We assume workers spend half day at work, other half mixing in general population. Alternatively, workers do 12hr shifts. Units are given where appropriate in the Value column.
[b] Manually set by matching the model output to the infection peak dates in the NHS Lothian data [14, 24].
[c] Initial model exploration indicated that higher rates (steeper drops in R_t) resulted in infection peaks (in general public and residents) falling too quickly compared to the NHS Lothian data [14, 24]. The value of 0.5 corresponds to a descent of ~ 2 weeks.
[d] We assume the reproductive rate for every sub-population drops to the Scottish government's [34] estimated R_t after lockdown (so $\omega^C_{low} = \omega^G_{low}$).
[e] Equals 0 to reflect the policy change to essential visitation only [14, 19], and to avoid the complication of modelling end-of-life visitation.

Data

NHS Lothian is the second-largest health board in Scotland [26], providing public health services to an estimated 907,580 people (2019 mid-year population estimate [27]). The daily confirmed positive tests of COVID-19 cases reported across

the entire health board were taken from the Public Health Scotland Open Data [24]. This data does not delineate which cases occurred in care homes, and thus, we retrieved the subset of cases in care homes from Burton et al. [14], which reports a 7-day average of confirmed cases in care home residents. Weekly COVID-19 deaths at the NHS Lothian health board level come from National Records Scotland [25]. Care home resident deaths are a subset of these and are published in [14]. Both death data are weekly counts of registered deaths where COVID-19 is mentioned on the death certificate (either as the underlying cause or as a contributory factor) [25].

Parameters set using evidence and assumptions

In this section, we describe our assumptions on parameters, parameters estimated from the literature, and parameters manually calibrated to the data.

A Scottish population study between 10th April to 15th June [31] estimated a combined adjusted seroprevalence across their study period (first wave = 10th April to 15th June) of 4.3% (95% CI 4.2%-4.5%). As of the week beginning 15th June 2020, there had been 18,077 positive tests [24], which as a percentage of Scotland's population (2019 census [27]) is $\sim 0.33\%$. We use this information to assume a constant reporting rate in the first wave for the general public of $r_G = 0.33/4.3 \sim 0.077$.

In Scotland, the policy from the start of March to 16th April 2020 was to test only the first few symptomatic care home residents, and afterwards, was to test all symptomatic residents [14]. Assuming when there is an outbreak in a home, 40% of the residents end up infected (40% incidence) [28, 29]. Given 48 residents per care home, until 16th of April we assume a reporting rate of (a few tested)/(total infected) $= 3/(0.4 \times 48) = 5/32$. After 16th April, we assume all the symptomatic cases are reported, giving a reporting rate of 4/5 (an estimated symptomatic proportion of COVID-19 cases in long term aged care is 80% [30]). Between the start of our simulation (6th March 2020) and 16th April 2020 is a time difference of 42 days, and between 17th April 2020 and the end of our simulation period (15th June 2020) is a time difference of 60 days. Therefore, for 42 days, we assume a reporting rate of 5/32, and for 60 days, it is 4/5. The weighted average and constant CH reporting rate over the simulation period is $r_C = (5/32)(42/102) + (4/5)(60/102) \sim 0.53$.

Until the 17th of April, we assume the staff reporting rate was the same as the general public (0.077). From then on, we assume the care home testing policy change (on the 17th of April) extended to their staff [14], and the reported percentage of cases was 83% (the symptomatic proportion [30]). Our weighted average and constant staff reporting rate over the simulation period is $r_S = 0.077 \times (42/102) + 0.8360/102 \sim 0.52$.

There are two constant death rates in our model: a resident death rate (μ_C) and a general population death rate (μ_G). We assume care home staff have the same death rate as the general population. There were ~ 899 positive tests and 423 deaths in NHS Lothian care home residents over the study period. Using our resident reporting rate, we estimate there were $899.1/0.53 \sim 1697$ total residents infected with COVID-19 over the study period. Therefore, we estimate a resident death rate of $\mu_C = 423/1697 \sim 0.25$. Similarly, there were 3123 total positive tests and 709 deaths over the study period in NHS Lothian overall. Using our general reporting rate, r_G, we estimate a general population death rate of $\mu_G = 709/(3123/0.077) \sim 0.017$.

Under our parameterisation, the timing of the drop in reproductive rates for care home residents (ω^C_{end}) and for the general population (ω^G_{end}) control the timing of peak infections in each respective population, independent of all other parameter values. This is linked to the reproductive rate function ($logi$) at an inflection point at $t = \omega^x_{end}$ (where $logi(\Omega_x)$ takes the value of $(\omega^x_{high} + \omega^x_{low})/2$). Therefore, we manually set these parameters (ω^C_{end}, ω^G_{end}) by matching the model output to the infection peak dates in the NHS Lothian data [14, 24].

The ω^G_{rate} and ω^C_{rate} parameters were assumed to be 0.5. Initial model exploration indicated that higher rates (steeper drops in R_t) resulted in infection peaks (in general public and residents) falling too quickly compared to the NHS Lothian data [14, 24]. From sensitivity analysis, we found that changing the values of these parameters does not affect the disease dynamics (in terms of total infections/deaths). The ω_{rate} parameter controls the steepness of the descent from ω_{high} to ω_{low}, however the timing of the start of the descent changes to almost cancel out the effect of changing the steepness of this drop. The value of $\omega^G_{rate} = \omega^C_{rate} = 0.5$, corresponds to a descent of $\sim$ 2 weeks. For ω^γ_{rate} we chose the value of 3 to follow the rapid visitation policy changes in care homes [14, 19].

The ω^G_{low} and ω^C_{low} values were set to 0.6, the estimated R_t after the first wave in Scotland [34]. Due to the uncertainty in the timing of the drop and the R_t peak value, we did not use this source for the ω^G_{high} and ω^G_{end} parameters.

We have made the simplifying assumption of $\omega^\gamma_{low} = 0$, to avoid the complications of modelling end-of-life visitation in care homes.

Closed environments are conducive to COVID-19 transmission and superspreading events [35], therefore we assume pre-lockdown transmission rates within care homes are not less than the general populations, $\omega^C_{high} \geqslant \omega^G_{high}$.

Care homes operate with staff under differing working hour schedules. This includes care homes having a day and night shift ($\delta = 0.5$), three 8 hour shifts ($\delta = 0.33$) or an uneven distribution of staff spread throughout the day. For simplicity, we assume that all homes operate under two 12 hour shifts per day, i.e., $\delta = 0.5$. Other shifts are explored in the sensitivity analysis.

We make a number of assumptions about the population initially infected. Workers were not initially infected in the model. In the general population, we assume an equal amount of exposed and infected individuals (with and without symptoms) i.e., $E_G(0) = I_G(0) + A_G(0)$. $H_{seeded} = |\{C_j \in \{C_1, \ldots, C_m\} : E_{C_j}(0) > 0\}|$ care homes were seeded with infections, representing introductions such as hospital discharges. To account for the delay in infections at the start of the pandemic in care homes compared to the general population, as seen in the data Fig 2, we assume for all $j \in \{1, \ldots, m\}$, $I_{C_j}(0) = A_{C_j}(0) = 0$. We seeded the homes so that initially infected homes lay equally spaced on the circle sharing structure (see Fig 1 (b)). If a home is seeded then we assume $E_{C_j}(0) = 1$, and if not, $E_{C_j}(0) = 0$.

Data fit

While some parameter values can be found based on the external data and literature, as shown in the previous section, other parameters were estimated using a formal fit to

the cases and deaths data for the Lothian NHS health board (Table 1). These parameters were free to vary subject to constraints based on a combination of assumptions and information from the literature. We used the method of least-squares, aggregating the error of model output against the four data sets for NHS Lothian cases and deaths and choosing the parameter set which minimised this error. The data for NHS Lothian population cases and care home cases were in the form of daily and seven day averages respectively. The death data for both the NHS Lothian population and care homes were in weekly counts. To make the fitting consistent, we transformed the daily and seven day average data into weekly data for conformity (Fig 2). The constraints on the parameters in our model, described in the previous section, left 6 free parameters for formal fitting. Their ranges used for the data fit are shown in Table 2.

Table 2. Parameters used for the data fit and the sets of values simulated over.

Parameter	Values considered
ω^C_{high}	$\{3.3, 3.4, \ldots, 5\}$
ω^G_{high}	$\{3.3, 3.4, \ldots, 4.5\}$
ε	$\{0.1, 0.2, \ldots, 0.5\}$
$E_G(0)$	$\{100, 110, \ldots, 180\}$
ω^γ_{high}	$\{0.042, 0.083, 0.17\}$
H_{seeded}	$\{1, 2, \ldots 10\}$

To investigate the question of how many care homes were exposed at the start of the pandemic, we ran the fitting separately for H_{seeded} fixed at 1 through 10. We simulated the model over 21,060 combinations of the remaining parameters to calculate the least-squares, for each value of H_{seeded}. We investigate the distribution of the parameters in Table 2 in the top ten best fitting scenarios, for each value of H_{seeded}.

Sensitivity analysis

After identifying the parameter set that minimises the least-squares, the base case (Table 2), we performed a sensitivity analysis. We measured the change in each population's deaths when shifting individual parameters in Table 3 from the base case. This allowed us to assess the relative impact of individual parameters on each population. We also measured the change in deaths in each population when changing pairs of the time-share parameters (ω^γ_{high}, δ, ε), keeping all other parameters at the base case. The results were stored in a 50×50 grid and visualised using heat-maps to determine the key modes of COVID-19 ingress into and throughout care homes.

Table 3. Parameters involved in the sensitivity analysis. Sensitivity shift is the unit of change used for each parameter from its base case. These values were chosen to measure the change in each population's deaths to small perturbations of individual parameters from it's base case.

Parameter	Sensitivity shift
ω^C_{end}	1
ω^C_{high}	0.1
ω^C_{low}	0.1
ω^C_{rate}	0.1
ω^G_{end}	1
ω^G_{high}	0.1
ω^G_{low}	0.1
ω^G_{rate}	0.1
δ	0.1
ε	0.05
ω^γ_{high}	0.0167
$E_G(0) = I_G(0) + A_G(0)$	10
λ	0.3
τ	0.4

Results

In this section we first show how the model captures the NHS Lothian data for cases and deaths in the period from March to June 2020, and subsequently show how sensitive the results are to changes in key parameters.

Data fit

The model captures the key features of the COVID-19 related cases and deaths in both care home and general populations, Fig 2. The minimum aggregate least-squares error was 33,042, with our model predicting 3,165 total cases and 817 total deaths compared to the total 3,123 cases and 709 deaths in the data. The average difference between data and predictions was 3.5 cases/deaths per week. In care homes we predicted 871 cases and 411 deaths compared to 903 cases and 423 deaths in the data. Our model does not predict the initial jump in deaths in care homes due to our assumption that infection reporting is constant. Further, our model overestimates the number of deaths for all populations despite a good fit for the cases, as the calculation of death rates is tied to the reporting rates.

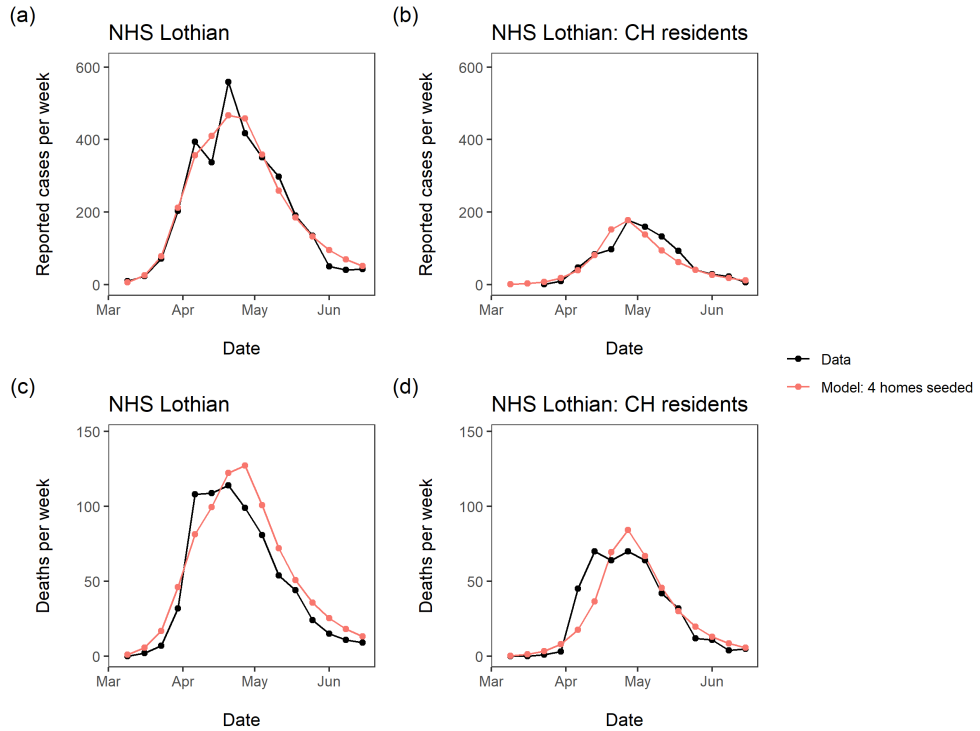

Fig 2. Surveillance data and fitted model. Data used for fitting are black lines, and model solution with parameter values in Table 1 are red lines. (a) reported cases per week for all NHS Lothian inhabitants (care home residents, workers and the general population); (b) reported cases per week in NHS Lothian care home residents; (c) deaths per week for all NHS Lothian inhabitants (care home residents, workers and the general population); (d) deaths per week in NHS Lothian care home residents.

The reproduction rates change rapidly over the period of April - May 2020, Fig 3 reflecting the delayed effect of the lockdown. The care home resident population's fall in reproduction rate is delayed by ~ 3 weeks compared to the fall in the general population. This delay is informed by the data, due to the ω_{end} parameters controlling the timing of the peaks in Fig 2 - we must be careful when attempting to interpret this delay.

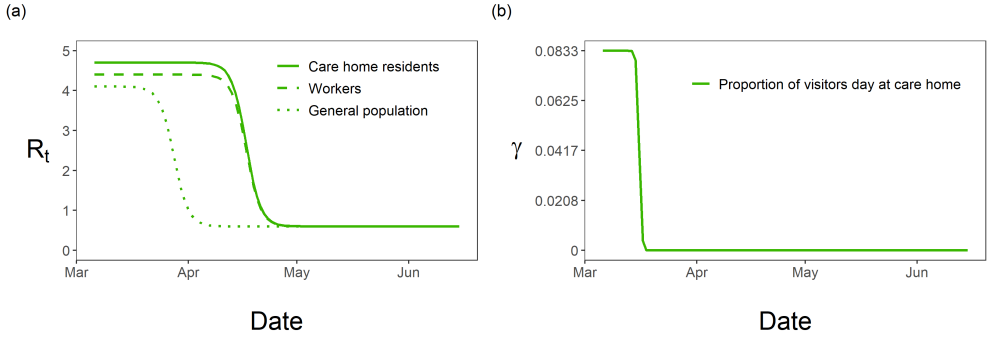

Fig 3. Fitted time-dependent parameters. (a) Fitted reproductive numbers over time for care home residents, $R_C(t)$, workers, $R_W(t)$, and general population, $R_G(t)$; (b) fitted visitation, γ, over time with drop highlighting the change in policy.

In order to assess the initial level of care home exposure to virus, we consider the quality of fit as a function of H_{seeded}. The minimum sum of squares of residuals takes the shape of a parabola, with a minimum at $H_{seeded} = 4$, see Fig 4. This suggests that a relatively small number of homes were initially exposed to COVID-19.

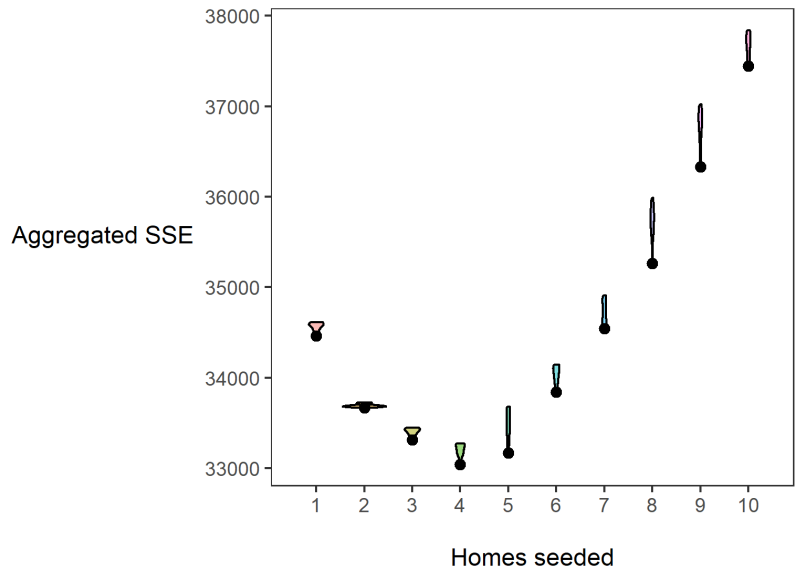

Fig 4. Quality of fit as a function of homes seeded. Each violin is the distribution of aggregated sum of squared errors (SSE) in the top ten best-fitting parameter sets, for a number of homes seeded. Black dots indicate the minimum aggregated SSE achieved for each home seeded.

The optimal choice (in terms of the least-squares criterion) for the parameters as used in the data fit (Table 2) is relatively stable with respect to changes in H_{seeded}, Fig 5. The pre-lockdown reproduction rate in care homes, ω^C_{high}, appears stable in the range of 4.5 to 4.7, changing for 10 homes seeded with the optimal value lowering to 3.9. This highlights the clear link between the reproduction rate and the exposure of care homes at the beginning of the pandemic. For the pre-lockdown reproduction rate in the general population, ω^G_{high}, we see a stable optimal value in the range of 3.9 to 4.2. We see a lot of uncertainty in the proportion of staff shared ε, for H_{seeded} outside of the range of 4 to 8. The distribution for ε changes for $H_{seeded} = 10$, with the optimal value going back up to 0.5. This coincides with the rise in $E_G(0)$ and the substantial fall in ω^C_{high}. This points to the correlation between these 3 parameters. Similarly, there is a lot of uncertainty in the value of pre-lockdown visitation, ω^γ_{high}, with an optimal choice for every value considered in our fitting as we vary H_{seeded}. This uncertainty highlights that the parameters in our model are highly correlated. As seen in Fig 6 and Fig 7, Fig 5 can also be seen to hint at how effective these parameters are at affecting the outcome of the model. For example, the variability in the chosen value of ε and ω^γ_{high} can also be attributed to the relatively small affect they have on the model outcome.

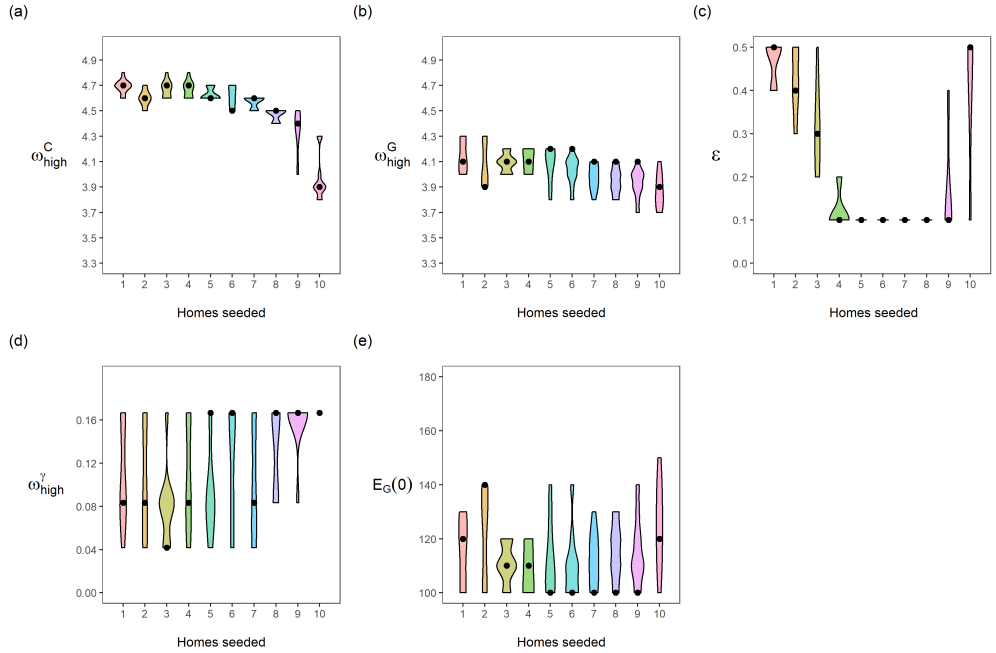

Fig 5. Distribution of fitted parameters as a function of homes seeded. Each panel is a different calibrated parameter. Each violin in a panel is the distribution of individual parameters in the top ten best fitting parameter sets, for a number of homes seeded. (a) pre-lockdown care home resident R_t, ω^C_{high}; (b) pre-lockdown general public R_t, ω^G_{high}; (c) staff sharing, ε; (d) visitation pre-lockdown, ω^γ_{high}; (e) general public seeded cases, $E_G(0) = I_G(0) + A_G(0)$. Black dots indicate the parameter value giving the lowest aggregated least square error, for each number of homes seeded.

Sensitivity analysis

Fig 6 indicates the sensitivity of the predicted deaths to the parameters in Table 3 for care home residents, workers and general population. Predictions are most sensitive to the infectious period, τ, and latency period, λ. The parameters $\{\omega^C_{end}, \omega^C_{high}, \omega^C_{low}\}$ significantly influence care home and worker deaths, without affecting predicted deaths in the general population. Interestingly, a change in one day from when the care home reproduction rate drops, results in almost 10% change in predicted resident and worker deaths. The parameters, $\{\omega^G_{end}, \omega^G_{high}, \omega^G_{low}\}$, controlling the timing of the reproduction rate and it's value before and after lockdown for the general population, significantly affect predicted deaths in the general population. Interestingly, changing the value of δ by 0.1 (20%) results in very little effect to the residents ($< 5\%$) and general population ($< 1\%$) but a 15% change in predicted worker deaths.

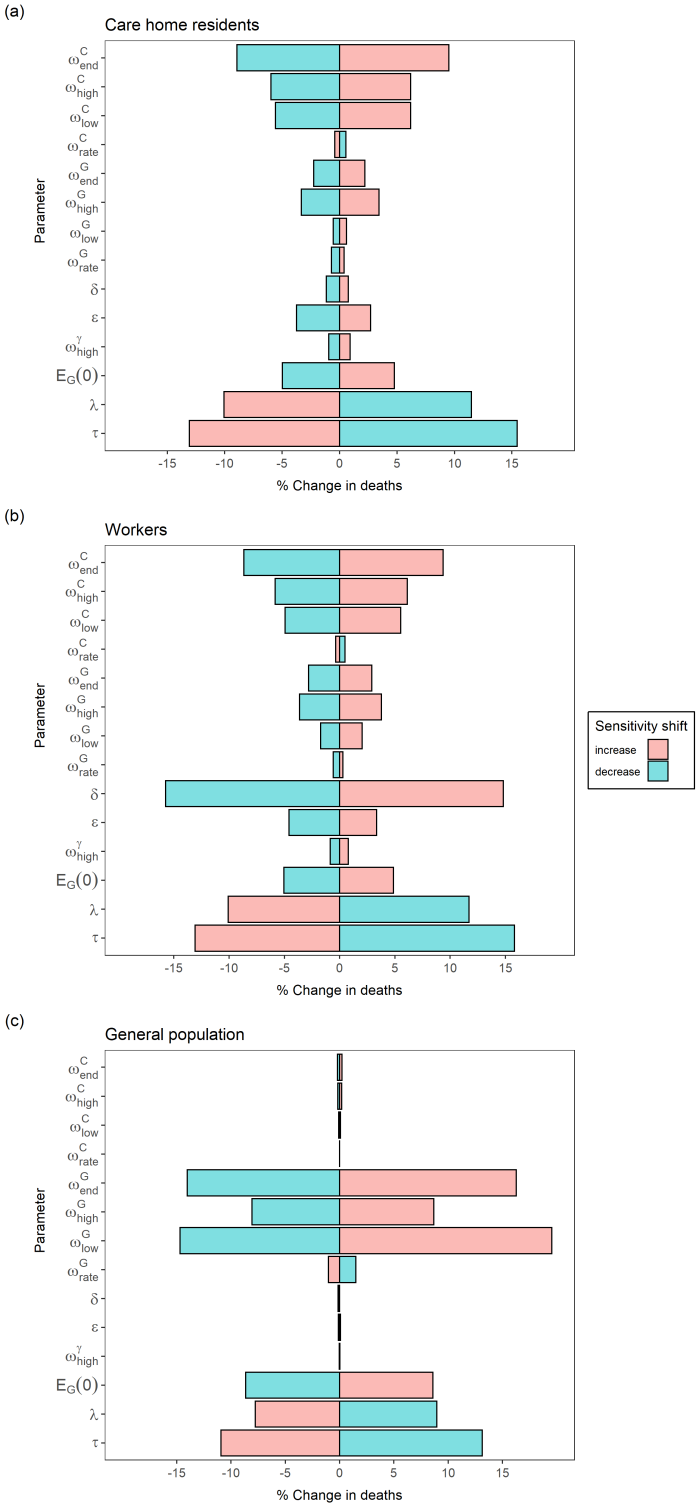

Fig 6. Sensitivity of the final deaths in each population to perturbations in model parameters. Each bar shows the % change in final deaths in a population caused by shifting an individual parameter from the base case, keeping all other parameters fixed at the base case (Table 1). Each parameter is increased or decreased from its base case value by the corresponding 'sensitivity shift' value in Table 3.

In our model, the path of staff catching infections from the community is controlled by δ; staff spreading infections between homes through staff sharing by ε; and visitors bringing infections into homes from the outside community by ω^{γ}_{high}. We investigate the relative risk of these COVID-19 pathways into care homes using Fig 7. This shows the combined impact of varying pairs of these parameters on the total resident deaths. Changing staff sharing, ε, and staff shift patterns, δ, the final number of predicted resident deaths do not change significantly, apart from when $\delta \sim 1$ (Fig 7). When considering $\delta = 1$, we should also restrict $\varepsilon = 0$, as staff living in care homes would not be shared across them. Varying ε and ω^{γ}_{high} does not significantly effect predicted deaths while holding all other parameters constant at their respective values in Table 1. Changing staff shift patterns (δ), has the largest impact on the predicted deaths, especially in the extremes. Reducing ε and ω^{γ}_{high} alone are weaker, but ε has a comparatively larger effect than ω^{γ}_{high}. Reducing ε to 0 and increasing δ to 1 together creates an even higher impact on the predicted deaths Fig 7, showing an outbreak severity reduction of $\sim 75\%$ compared to the observed values in Lothian during the first wave. Reducing pre-lockdown visitation from 2 visiting hours per resident to 0 hours per resident, would reduce our predicted first-wave deaths by 20 to 40, about 10% of the death count in the first wave in NHS Lothian.

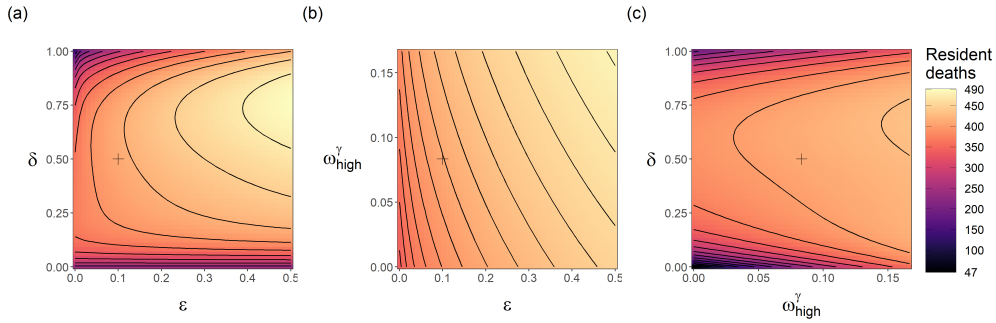

Fig 7. Sensitivity of the final resident deaths to the time-share/mixing parameters ($\delta, \varepsilon, \omega^{\gamma}_{high}$). Proportion of CH staff at work is δ, proportion of staff shared between homes is ε, and pre-lockdown visitation is ω^{γ}_{high}. Each panel shows the combined impact of varying two of the time-share/mixing parameters, with all other model parameters fixed as the base case (Table 2). The black lines in each panel are isoclines. The cross in each panel indicates the base case value for each parameter.

Discussion

To identify and rank the key modes of ingress into care homes, we used a combination of modelling, data fit and simulations. We find that homes in our model are more at risk to outbreaks through staff infections from the general population, relative to visitation or staff sharing. We also find that outbreaks in our model were not significantly driven by hospital discharges. These findings coincide with the results from Rosello et al., who used a stochastic compartmental model on single care homes in England [9]. We additionally find that in our model the drop in within-home reproduction rate was 3 weeks behind that of the general population.

Changing worker shift patterns in care homes (δ) only weakly affects the model outcome for most "reasonable" values e.g. a 2-shift pattern ($\delta = 0.5$) or 3-shift pattern ($\delta = 0.33$). It is only in the extremes where substantial differences are seen. δ close to 1, alongside low pre-lockdown visitation, greatly reduces the final outbreak size in care

homes. Thus, our results point to a strategy of staff living-in with residents, in conjunction with timely lowering of visitation, as an effective pandemic response. This happened in France, where care home outbreak size was reduced significantly in care homes where staff self-confined [36]. If living within the care home is not possible, this result of very high levels of δ may imply that the strategy of segmenting the staff away from both care home residents and the general population whilst they are not at work would be effective, e.g., organisation of accommodation for care home workers [3]. From our model, we see the most effective solution to keeping care homes safe from infection is to focus on the pathway from general population to workers to residents.

Eliminating staff sharing did not eliminate outbreaks in our model simulations, suggesting that this was not the primary route of infection entering homes. Supporting the literature [3, 20, 37], reducing staff sharing does reduce the outbreak severity; in our model, this impact is low compared to other routes. This conclusion is limited due to our assumption of the circle sharing contact structure, which in turn reflects limited data availability regarding the contact structure of the care home industry in Lothian (due to commercial sensitivity). A different contact structure could result in staff sharing leading to more/less exportation of infection from homes with outbreaks. A more thorough examination on the contact-structure of this system and how that impacts disease spread dynamics would be an important contribution to the literature. One way to achieve this would be to consider an addition of highly connected hubs [38]. We expect that including highly connected worker populations will increase the effect of staff sharing. In our simulations, staff sharing has an effect when there exists a non-uniform distribution of infections in worker sub-populations. Worker sub-populations acting as hubs would acquire disease quicker and skew the distribution of infection amongst worker sub-populations. However, the general population strongly connects all nodes in the network, and dominates the impact of the staff sharing network on disease spread. This is because we assume a single general population with full mixing. At the geographical scale we model (a health board) this is an appropriate assumption, although it would not hold for larger scales of heterogeneity, e.g., the national level (Scotland).

A reduction in visitation reduces predicted resident deaths, as speculated in [3]. Our model predictions support findings that visitation to care homes was not the driving cause of infection in care homes [39]. Since visitation was banned, the evidence for visitation causing outbreaks is limited. Investigation with constant visitation would be necessary to see how the outcome would be different if visitation did not change at all; this was not the focus of our investigation.

Our parameter estimations suggest that after the nation-wide lock down, R_t within care homes dropped three weeks after it did in the wider population. Several possible explanations exist, including differences in testing availability or testing strategies and the difficulty in controlling care home resident interactions to lower disease transmission. Dropping care homes' reproduction rate 1 day earlier, results in a 10% reduction in resident and worker deaths. Therefore a more thorough investigation into the delay between the general populations' and care homes' reproduction rate is necessary to aid closing this gap in future outbreaks.

From the data fit, there were a low number of care homes infected at the beginning of the first wave ($H_{seeded} = 4$). These initial infections could represent hospital discharges or any other pathway. The result supports the claim from a report on English care homes that resident discharges from hospitals were not the primary cause

of care home outbreaks [40]. The hospital discharges are only included as initial care home infections, H_{seeded}. In reality, discharges continued during the first wave [14] and more detailed data will be needed to address this problem.

We made a number of simplifying assumptions. Our model does not explicitly account for the variation in susceptibility with age [41]. It is only implicitly addressed by considering different values of β within and outwith of care homes, while keeping the staff and general population homogeneous. Due to the unavailability of data regarding care home worker infections, we expressed worker transmission rates in terms of transmission rates for care home residents and the general population. We assumed the resident-resident and resident-worker transmission rates were equal. However, contacts between care giving staff and residents are likely more frequent and closer than between residents. On the other hand there may be more adherence or better knowledge of how to use PPE among staff. Also, contact between residents could be reduced more easily during the pandemic [3].

The data fit was achieved by minimising the aggregated sum of squared error for each of the four time-series. This method requires the errors to be independent, follow a normal distribution and for the variance to be constant. With the data, we could not estimate the variance over time. To mitigate the effect of the differing variance of the data sets, we shifted the four time-series to the same scale, this being weekly cases/deaths.

The size of individual care homes is believed to be the main factor that influences the likelihood of a care home outbreak [14, 20, 42]. Larger homes typically have more staff and therefore a higher chance of experiencing an outbreak before the smaller ones. In general, we expect larger care homes to receive an increased force of infection from all sources, proportional to its increased size, and therefore an increased outbreak risk. This in turn could increase risk for smaller homes directly connected to the larger ones through staff sharing and visitations, and the overall outbreak risk. However, this effect could balanced by a lowered risk associated with small care homes, with the total population size kept constant. In our model, we assume a uniform home size in order to keep the model generic. As a result (and since the model is deterministic) the risk of staff and visitors bringing in infections is the same for all care homes, which may result in underestimation of the initial rate of spread. An obvious extension of the paper would be to consider various sources of heterogeneity, including size.

The National Records Scotland death data used were the dates of death registration, not the date of death. This is limiting, as we are an average of 3 days behind in the prediction of deaths [25]. The data for care home resident deaths includes deaths in hospital; including nosocomial infections, which we do not take into account into our model. We expect this not to limit the interpretation of our results, considering hospital deaths of care home residents make up approximately 5% of the total care home resident deaths [25].

We do not distinguish explicitly between symptomatic and asymptomatic individuals, as in (**A**) there are people with symptoms that would have been missed by testing. However, asymptomatic infections implicitly affect this model's reporting rates. We do not explicitly model self-isolation or any behavioural change after infection, nor delays or changes in reporting. For simplicity, reported infected individuals are reported immediately. Since we are not explicitly modelling behaviour

change once infected, we do not expect incorporating reporting delays to significantly affect our results on the infection pathways. Reporting differed over time, especially in the early weeks of the pandemic when testing was scarce; in care homes, the national policy was to test only the first few symptomatic residents [14]. With our constant reporting rate, we overestimate the number of positive cases prior to the policy change, and underestimate the cases afterwards. A time-dependent reporting rate would impact our death rate and reproductive rate parameters and therefore requires further study.

Data regarding care home outbreaks were limited due to the commercial nature of care home organisations in Scotland. Making this data available would allow for additional modelling approaches. Adding in further heterogeneity into the system by including a distribution of home sizes and types would further improve the modelling approach. Including a stochastic component to this model could lead to more insight into "super-spreader" events in care homes [35] and their effect on epidemic response. Finally, this work focuses on disease dynamics over the course of the first wave, it would be interesting to model this system for future waves and to incorporate the implementation of vaccinations. Extending this system to the rest of Scotland or even the entirety of the UK would be a clear extension of this work.

Conclusion

In our model, the primary route of introduction to the care home sector is from the staff via the general population. Our results also agree with other findings in the literature which suggest that visitation was not the major cause of COVID-19 establishment in care homes [39] and that hospital discharges accounted for a small number of care home outbreaks early on in the pandemic [40]. Prioritising the protection and monitoring of staff, second to timely reductions in visitation and staff sharing, are what our model suggests for an effective reduction in outbreak size. Our findings also show difficulties in halting the spread of COVID-19 within the homes in tandem with the general population. This highlights the need for more planning and support for care homes and their staff in organising quick and effective responses to emerging pandemics. The lessons to be learned from the first wave of COVID-19 will be crucial to minimise the damage from future outbreaks.

Acknowledgements

The authors thank Public Health Scotland for their support, and Prof. Bruce Guthrie (University of Edinburgh) for sharing the Lothian care home data. EMT has been partially supported by the University of Strathclyde Student Excellence Award and MB by Defra. Results were obtained using the ARCHIE-WeSt High Performance Computer (www.archie-west.ac.uk) based at the University of Strathclyde. Results were also obtained using the Pin cluster at the University of Stirling.

References

1. World Health Organisation. WHO Coronavirus (COVID-19) Dashboard; 2021. Available from: `https://covid19.who.int/`.

2. Bell D, Comas-Herrera A, Henderson D, Jones S, Lemmon E, Moro M, et al. COVID-19 mortality and long-term care: a UK comparison. LTCcovid; 2020.

Available from: https://ltccovid.org/wp-content/uploads/2020/08/COVID-19-mortality-in-long-term-care-final-Sat-29-1.pdf.

3. Social Care Working Group (SCWG) for the Scientific Advisory Group for Emergencies. Commission: What are the appropriate layers of mitigation to deploy for care homes in the context of post vaccination risk landscape? UK Government; 2021. Available from: https://www.gov.uk/government/publications/scwg-what-are-the-appropriate-mitigations-to-deploy-in-care-homes-in-the-

4. Nguyen LKN, Megiddo I, Howick S. Challenges of infection prevention and control in Scottish long-term care facilities. Infection Control & Hospital Epidemiology. 2020;41(8):943–945. doi:10.1017/ICE.2020.113.

5. Smith DRM, Duval A, Pouwels KB, Guillemot D, Fernandes J, Huynh BT, et al. Optimizing COVID-19 surveillance in long-term care facilities: a modelling study. BMC Medicine. 2020;18:1–16. doi:10.1186/S12916-020-01866-6.

6. Nguyen LKN, Howick S, McLafferty D, Anderson GH, Pravinkumar SJ, Van Der Meer R, et al. Evaluating intervention strategies in controlling COVID-19 spread in care homes: An agent-based model. Infection Control and Hospital Epidemiology. 2020; p. 1–11. doi:10.1017/ice.2020.1369.

7. Nguyen LKN, Megiddo I, Howick S. REPORT 3: IMPACT OF VARIOUS VACCINATION COVERAGES ON THE SPREAD OF COVID-19 AND DEATHS IN CARE HOMES. LTCcovid; 2021. Available from: https://ltccovid.org/wp-content/uploads/2020/12/Report-3_Care-homes_Vaccination_Strathclyde.pdf.

8. Nguyen LKN, Howick S, McLafferty D, Anderson GH, Pravinkumar SJ, Van Der Meer R, et al. Impact of visitation and cohorting policies to shield residents from covid-19 spread in care homes: an agent-based model. American Journal of Infection Control. 2021;doi:10.1016/j.ajic.2021.07.001.

9. Roselló A, Barnard RC, Smith DRM, Evans S, Grimm F, Davies NG, et al. Impact of non-pharmaceutical interventions on SARS-CoV-2 outbreaks in English care homes: a modelling study. medRxiv [Preprint]; 2021. medRxiv 2021.05.17.21257315. Available from: https://www.medrxiv.org/content/early/2021/05/18/2021.05.17.21257315.

10. Overton CE, Stage HB, Ahmad S, Curran-Sebastian J, Dark P, Das R, et al. Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. Infectious Disease Modelling. 2020;5:409–441. doi:10.1016/j.idm.2020.06.008.

11. Oberhammer J. Social-distancing effectiveness tracking of the COVID-19 hotspot Stockholm. medRxiv [Preprint]; 2020. medRxiv 2020.06.30.20143487. Available from: https://www.medrxiv.org/content/10.1101/2020.06.30.20143487v2.

12. Bunnik BADv, Morgan ALK, Bessell PR, Calder-Gerver G, Zhang F, Haynes S, et al. Segmentation and shielding of the most vulnerable members of the population as elements of an exit strategy from COVID-19 lockdown. Philosophical Transactions of the Royal Society B. 2021;376(1829). doi:10.1098/RSTB.2020.0275.

13. Public Health Scotland. Discharges from NHS Scotland Hospitals to Care Homes between 1 March and 31 May. Public Health Scotland; 2020. Available from: https://publichealthscotland.scot/publications/discharges-from-nhsscotland-hospitals-to-care-homes/discharges-from-nhsscotland-hospitals-to-care-homes-between-1-march-and-3

14. Burton JK, Bayne G, Evans C, Garbe F, Gorman D, Honhold N, et al. Evolution and effects of COVID-19 outbreaks in care homes: a population analysis in 189 care homes in one geographical region of the UK. The Lancet Healthy Longevity. 2020;1(1):E21–E31. doi:10.1016/s2666-7568(20)30012-x.

15. Knock ES, Whittles LK, Lees JA, Perez-Guzman PN, Verity R, FitzJohn RG, et al. Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. Science Translational Medicine. 2021;13(602). doi:10.1126/scitranslmed.abg4262.

16. McKeigue PM, Colhoun HM. Evaluation of "stratify and shield" as a policy option for ending the COVID-19 lockdown in the UK. medRxiv [Preprint]; 2020. medRxiv 2020.04.25.20079913. Available from: https://www.medrxiv.org/content/early/2020/04/30/2020.04.25.20079913.

17. Ripperger TJ, Uhrlaub JL, Watanabe M, Wong R, Castaneda Y, Pizzato HA, et al. Orthogonal SARS-CoV-2 Serological Assays Enable Surveillance of Low-Prevalence Communities and Reveal Durable Humoral Immunity. Immunity. 2020;53(5):925–933. doi:10.1016/j.immuni.2020.10.004.

18. Dan JM, Mateus J, Kato Y, Hastie KM, Yu ED, Faliti CE, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. Science. 2021;371(6529). doi:10.1126/science.abf4063.

19. Cabinet Secretary for Health and Sport. Coronavirus (COVID-19) social care response - 13 March 2020. The Scottish Government; 2020. Available from: https://www.careinspectorate.com/images/COVID-19_-_Letter_from_Cabinet_Secretary_for_Health_and_Sport_-_Social_care_guidance_-_13_March_2020.pdf.

20. Reilly J, Crawford D, Boyle DO. CARE HOME REVIEW: A rapid review of factors relevant to the management of COVID-19 in the care home environment in Scotland. The Scottish Government: Cabinet Secretary for Health and Sport; 2020. Available from: https://www.gov.scot/publications/root-cause-analysis-care-home-outbreaks/.

21. Office For National Statistics. Impact of coronavirus in care homes in England: 26 May to 19 June 2020. Office For National Statistics; 2020. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/impactofcoronavirusincarehomesinenglandvivaldi/26mayto19june2020.

22. Rădulescu A, Williams C, Cavanagh K. Management strategies in a SEIR-type model of COVID 19 community spread. Scientific Reports. 2020;10(1). doi:10.1038/s41598-020-77628-4.

23. Calvetti D, Hoover AP, Rose J, Somersalo E. Metapopulation Network Models for Understanding, Predicting, and Managing the Coronavirus Disease COVID-19. Frontiers in Physics. 2020;8. doi:10.3389/fphy.2020.00261.

24. Public Health Scotland. Scottish Health and Social Care Open Data: Daily Case Trends By Health Board; 2020. Available from: https://www.opendata.nhs.scot/dataset/covid-19-in-scotland/resource/2dd8534b-0a6f-4744-9253-9565d62f96c2.

25. National Records of Scotland. Deaths involving coronavirus (COVID-19) in Scotland; 2020. Available from: 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.

26. NHS Lothian. NHS Lothian: Our Services; 2021. Available from: https://services.nhslothian.scot/Pages/default.aspx.

27. National Records of Scotland. Population Estimates Time Series Data; 2019. Available from: https://www.nrscotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/population/population-estimates/mid-year-population-estimates/population-estimates-time-series-data.

28. Ladhani SN, Chow JY, Janarthanan R, Fok J, Crawley-Boevey E, Vusirikala A, et al. Investigation of SARS-CoV-2 outbreaks in six care homes in London, April 2020. EClinicalMedicine. 2020;26. doi:10.1016/j.eclinm.2020.100533.

29. Graham NSN, Junghans C, Downes R, Sendall C, Lai H, McKirdy A, et al. SARS-CoV-2 infection, clinical features and outcome of COVID-19 in United Kingdom nursing homes. Journal of Infection. 2020;81(3):411–419. doi:https://doi.org/10.1016/j.jinf.2020.05.073.

30. Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: Systematic review and meta-analysis. Journal of the Association of Medical Microbiology and Infectious Disease Canada. 2020;5(4):223–234. doi:10.3138/jammi-2020-0030.

31. Dickson E, Palmateer NE, Murray J, Robertson C, Waugh C, Wallace LA, et al. Enhanced surveillance of COVID-19 in Scotland: population-based seroprevalence surveillance for SARS-CoV-2 during the first wave of the epidemic. Public Health. 2021;190:132–134. doi:https://doi.org/10.1016/j.puhe.2020.11.014.

32. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nature Medicine. 2020;26:672–675. doi:10.1038/s41591-020-0869-5.

33. McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F, et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. BMJ Open. 2020;10. doi:10.1136/bmjopen-2020-039652.

34. The Scottish Government. Coronavirus (COVID-19): modelling the epidemic in Scotland (Issue No. 29). The Scottish Government; 2020. Available from: https://www.gov.scot/publications/coronavirus-covid-19-modelling-epidemic-issue-no-29/.

35. Majra D, Benson J, Pitts J, Stebbing J. SARS-CoV-2 (COVID-19) superspreader events. Journal of Infection. 2021;82(1):36–40. doi:https://doi.org/10.1016/j.jinf.2020.11.021.

36. Belmin J, Um-Din N, Donadio C, Magri M, Nghiem QD, Oquendo B, et al. Coronavirus Disease 2019 Outcomes in French Nursing Homes That Implemented Staff Confinement With Residents. JAMA Network Open. 2020;3(8). doi:10.1001/jamanetworkopen.2020.17533.

37. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, et al. Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. New England Journal of Medicine. 2020;382:2005–2011. doi:10.1056/NEJMoa2005412.

38. Barabási A, Bonabeau E. Scale-Free Networks. Scientific American. 2003;288;60–69. doi:10.2307/26060284.

39. Comas-herrera A. Rapid review of the evidence on impacts of visiting policies in care homes during the COVID-19 pandemic. LTCcovid; 2020. Available from: `https://ltccovid.org/wp-content/uploads/2020/11/Rapid-review-of-evidence-on-impacts-of-visiting-policies-in-care-homes-du pdf`.

40. Public Health England. A data linkage approach to assessing the contribution of hospital-associated SARS-CoV-2 infection to care home outbreaks in England, 30 January to 12 October 2020. UK Government; 2021. Available from: `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/983349/Data_linkage_approach_to_assessing_the_contribution_of_hospital-associated_SARS-CoV-2_infection_to_care_home_outbreaks_in_England.pdf`.

41. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Pearson CAB, et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. Nature Medicine. 2020;26:1205–1211. doi:10.1038/s41591-020-0962-9.

42. Burton JK, McMinn M, Vaughan JE, Fleuriot J, Guthrie B. Care-home outbreaks of COVID-19 in Scotland March to May 2020: National linked data cohort analysis. Age and Ageing. 2021;doi:10.1093/ageing/afab099.

Appendices

S1 Appendix

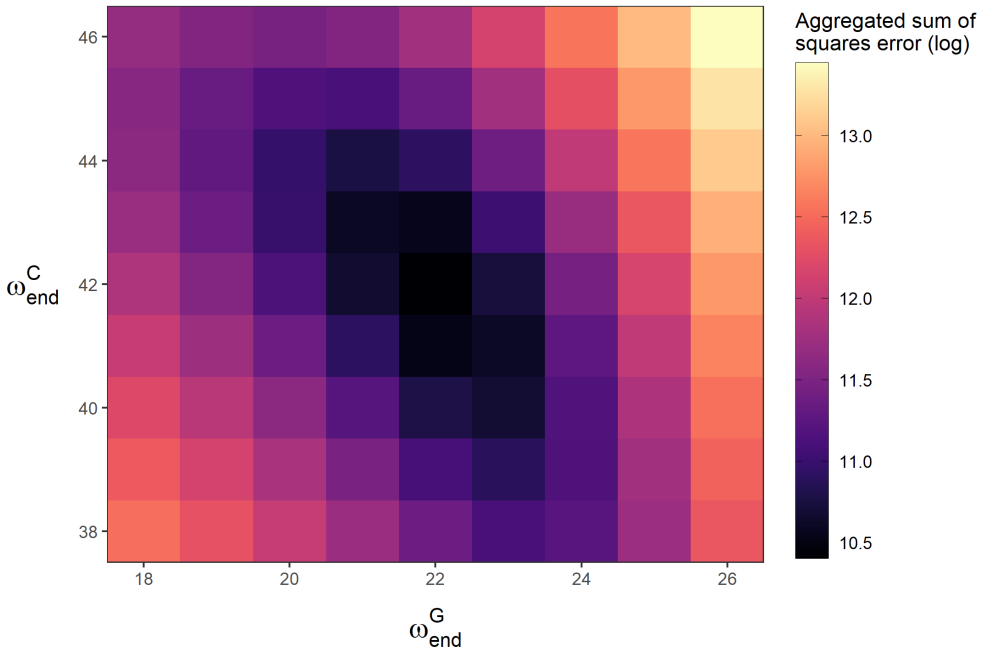

Fig 8. Natural log of aggregated sum of squared error in a $\omega^C_{end} - \omega^G_{end}$ parameter space. ω^C_{end} controls the timing of the drop in resident R_t, and ω^G_{end} controls the timing of the drop in general population R_t. These parameters were fitted manually, achieving a minimum for the values shown in Table 1. In the plot, all other parameters are held at the base case (Table 1).