Towards a behavioural model of Covid-19 spread

Umberto Gostoli[1] and Eric Silverman[1]

[1]MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, 99 Berkeley St., Glasgow G3 7HR
eric.silverman@glasgow.ac.uk

Abstract

In the past year, various models of Covid-19 spread have been proposed. While most of these models focused on the replication of the interactions' processes through which the virus is passed on from infected agent to susceptible ones, less effort has been devoted to the process through which agents modify their behaviour as they adapt to the risks posed by the pandemic. Understanding the way agents respond to Covid-19 spread is important as this behavioral response affects the dynamics of virus spread by modifying interaction patterns. In this paper, we present an agent-based model which includes a behavioural module determining the households' isolation propensity in order to understand the role of various behavioural parameters in the spread of Covid-19.

Introduction

Following its appearance in late 2019, the SARS-CoV-2 virus or Covid-19 has become the most significant global pandemic since the 1918 Spanish Flu. As of 28 April 2021 the World Health Organization records over 148 million confirmed Covid-19 cases worldwide, and over 3.1 million deaths (`https://covid19.who.int/`). The scale and severity of this global crisis, and the subsequent need for severe public health restrictions to be implemented worldwide, has generated significant interest in using agent-based modelling (ABM) to simulate the spread of the pandemic and its effects.

However, relatively few Covid-19 models have examined the impact of behavioural factors on how individuals respond to the risks posed by the pandemic; instead, much of the extant modelling work is highly abstracted, and uses very simplified behavioural models. For example, Wilder et al. (2020) proposed an agent-based model in which out-of-household contacts are based simply on a country-specific contact matrix containing the mean number of daily contacts agents of an age group have with agents from each of the other age groups. Similarly, Silva et al. (2020) present a model which allows agents to have differential exposure to risk according to their economic status, but does not include a facility for agents to modify their behaviour in accordance with their own perception of risk.

A few empirical studies have investigated the factors associated with adherence to self-isolation and lockdown policies. Smith et al. (2020) show that the level of compliance to lockdown measures depends on the perceived risk posed by the virus and the perceived social norms. Pullano et al. (2020) find that, in France, the reduced mobility after the lockdown was positively associated with the number of hospitalizations (which can be thought as a proxy for the risk posed by the virus) and the socio-economic conditions.

In this paper, we present an agent-based model of Covid-19 spread which accounts for the feedback process by which pandemic dynamics influence agent behaviour and, in turn, agent behaviour influences pandemic dynamics. The central element of this process is a behavioural model through which households reduce their level of social interaction as a response to the perceived risks posed by the virus. At this stage, our main motivation is the proposal of a proof-of-concept framework. We propose that simulations including these behavioural elements will potentially help policy-makers to design more effective interventions in future global health crises.

General framework

The simulation unfolds in two stages: in the first stage a demographic process proceeds in one-year steps, from year 1860 to year 2020, to create a population whose demographic and socio-economic characteristics roughly replicates those of the UK population. In the second stage, the Covid-19 spread is simulated through one-day steps, starting at the beginning of the year 2020 for 180 days.

Demographic process

This stage begins with the generation of an initial population of couples, which are randomly distributed on a 8×12-cell grid approximating the geography of the United Kingdom. Each cell of the grid represents a town, and within each town a number of houses is created proportional to the UK population density. Each year, a series of demographic events drive the population's dynamics: births, marriages, divorces and deaths. Empirical population data (in the form of UK

Census data) is integrated into the model's demographic processes in 1951.

Agent life-course With a certain age-specific probability the couple's female will give birth. Agents enter adulthood at the working age of 16: at this point they can either start working or continue in education, a choice that is repeated at two-year intervals, until the age of 24[1]. After education, agents become employed, taking a salary which is a function of the education level they have reached (which is a stochastic function of their parents' socio-economic status. See the *Socioeconomic status* subsection below). When agents reach retirement age (set at 65 in these simulations), they retire from employment.

Partnership Formation and Dissolution Once they reach working age, agents can form partnerships. Agents are paired randomly with probabilities that depend inversely on the agents' geographical distance from one another, their age and socioeconomic differences. Age-specific annual divorce probabilities determine whether a couple dissolves their partnership in each year.

Internal migration Relocation happens most frequently due to agents finding a partner in a different town. Male agents will also relocate to new houses once a partnership dissolves, and any children produced by that partnership stay with the mother. Retired agents with care needs may move in with one of their their adult children, with a probability determined by the their care need level and the amount of care supply in their child's household. Orphaned children are adopted by a household in their kinship network, or by a random family if there are no available households in their kinship network. Apart from these specific, event-driven cases, households also relocate to another town with a certain probability which is inversely proportional to the relative cost of relocation, defined as the ratio between the total cost of relocation and the households' per capita income.

Socioeconomic status and income Agents are placed in one of five socioeconomic status groups (SES groups), based on the Approximated Social Grade from the Office for National Statistics.

Every employed agent receives an hourly salary which is a function of its SES and its cumulative work experience. On the basis of the total income of the household's members, each household is assigned an *income quintile*, which affects a range of processes during the pandemic stage, such as the size of social networks, the probabilities to visit certain venues, the isolation propensities and the probabilities to develop different conditions (see the next subsection for details).

[1]These two-year intervals represent educational stages corresponding roughly to UK education levels: A-level, Higher National Diploma, Degree and Higher Degree.

The interaction processes

In the second stage, the Covid-19 spread is simulated from year 2020 for a period 180 days, through one-day time steps. We assume that, initially, the virus is brought into the UK from abroad, by international travellers, a process that we will call *exogenous infection*. Then, the virus spreads within the UK by means of two main spreading mechanisms: through the *social interactions* and through the *within-household interaction*. We assume that the social interaction process depends on the households' decisions to isolate. After the incubation period, the infected individuals develop various conditions and, at the end of the infection period, either recover or die.

Exogenous infections. In each period, a fixed share of the adult, susceptible population is infected exogeneously, i.e. through contacts they have with people abroad. We assume that the probability of being part of this group, which can be thought as composed of international travellers, depends on the agents' socio-economic status and the dimension of the town where they come from.

Domestic interaction. The domestic infection is part, with the infection through social interaction, of the endogenous spread of the virus. We assume that the probability that a susceptible agent is infected by a member of its household is an increasing function of the number of infected household's members and depends on the knowledge of the infection (as we assume that if an agent is knowingly infected the household member will adopt a prudential behaviour reducing the probability of transmission).

Social interaction. In our model, social interaction takes place in a series of venues the agents attend. At the beginning of the pandemic stage, a number of venues is created in each town proportional to the town's population. In each day, we allocate agents to the venues in their town in a way that is consistent with age-specific interactions' matrix for the UK, as estimated by Prem et al. (2017). Moreover, we assume that a certain percentage of agents visits other towns in each period, to account for daily intra-urban commuting.

Starting from empty venues, agents are sequentially allocated to venues in their town, with a probability which depends on the geographical distance between the agents' house and the venues' location and the difference between the mean of the socio-economic status of the venues' attendants and the socio-economic status of the agents to be allocated (the assumption being that the choice of venues is partly driven by socioeconomic affinity). The venues' attendants are sampled randomly from the population, so in each period, an agent can visit more than one venue.

The probability of an agent attending a venue being infected in proportional to the number of infected agents attending that venue.

The behavioural module

As the pandemic unfolds, the number of people attending the venues decreases as people become aware that the interaction with other people carries the risk for them of getting the virus. The prudential behaviour of the agents is represented by the *isolation rate*, R, a function of the venue's crowding which can take values from 0 (no isolation) to 1 (complete isolation). We assume that the agents' R is determined at the household's level, as the isolation of each household member affects the efficacy of the other members' prudential behaviour.

Loosely following the Epstein's 'Agent_Zero' cognitive architecture (Epstein, 2014), the household's R is affected by three factors: the public information regarding the virus' contagiousness and virulence (the *rational* factor); the occurrence of virus-related events (infections, hospitalizations, intubations or deaths) within the household's neighbors (the *emotional* factor); and the mean R of the household's neighbors (the *social* factor).

The role of public information. As for the rational part, the household's R, for a venue with n attendants, is given by:

$$R_n = 1 - e^{-\beta I L_n} \tag{1}$$

where I is the household's income quintile (from 1, the poorest, to 5, the wealthiest), L_n is the *expected loss* associated with visiting a venue with n attendants and β is a behavioural parameter. We assume, therefore, that the higher the household's income quintile the higher its isolation rate for a given expected cost, an assumption based on the empirical observation that the higher the income the greater the people's possibility to isolate. Moreover, a higher income means also a lower cost, in relative terms.

Following the behavioural function used by Green et al. (1999) to describe the discounted value of probabilistic gains (and losses), we define L_n as:

$$L_n = \frac{C}{(1 + \frac{h}{e^{\alpha A M_n - 1}})^s} \tag{2}$$

where C represents the expected cost associated with infection, given by:

$$C = p_H H + p_V V + p_D D \tag{3}$$

where the $p's$ are the age-specific public probabilities, in order, to be hospitalized, to be intubated and to die, if infected, and H, V and D are three parameters representing the costs associated with each of the three events (for simplicity, we set $H = 1$, $V = kH$ and $D = kV$, with k being a parameter greater than 1).

The denominator of 2, represents the *discounting factor* of the expected cost of infection: it is a decreasing function of the *perceived* probability of infection associated with attending a venue with n attendants, as a higher infection probability means a less discounted (i.e. higher) expected cost. While this probability cannot be but a very uncertain and subjective estimate, we can assume that it is somewhat associated with the virus' speed of circulation and the number of people attending the venue. In particular, we assume that the virus circulation is approximated by the *number of new infections*, A, and that the effect of the venue's crowdedness, M_n, grows with the number of attendants n (at a decreasing speed).

The perceived probability of infection, is also affected by two behavioural parameters, h and s, and a household-specific variable, α, which regulates the effect of A and M_n on the perceived probability of infection, variable that changes through the emotional effect of the observation of infections' events (see section below).

The emotional factor. Each household is associated with a network of *neighbors*. A household's neighborhood is the set of other households in which at least one member has a relationship, as a next-of-kin or as a friend, with at least one member of the household. This neighborhood network is created through the kinship relations between agents, which are the results of the demographic process, and by sampling a set of friends for each agent, with the probability of establishing a friendship depending on geographical locations and the differences between the agents' ages and socioeconomic statuses. The links between an household and its neighbors are associated with a weight which is the inverse of the geographical distance between the households' houses.

The role of the neighbors' network on the determination of an household's isolation rate is twofold: on one hand, the neighbors' mean isolation rate represents a 'point of attraction' for the isolation rate of the household (see next section); on the other, pandemic events happening within an household's network generate an *emotional bias* of the information through which the household's isolation rate is determined, that is, the probabilities of infection and of developing various conditions if infected. The assumption is that when agents observe pandemic events within their relatives and friends, these tend to have a disproportionate impact on the agents' assessment of the events' probabilities and, therefore, on their decision to isolate (i.e. agents tend to 'overreact').

The pandemic events the agents can observe are the infection, hospitalization, intubation and death of a member of an household belonging to their neighbors' network. Following Epstein (2014), we model the effect of these events on probabilities by means of a simplified version of the Rescorla-Wagner model (Rescorla, 1972). In general, when an agent observes a certain event, the new estimate of the probability for that event, p_{t+1}, will be given by:

$$p_{t+1} = p_t + \gamma(1 - p_t) \quad (4)$$

where γ is the increase rate of the estimate, a variable which depends on the number of events observed and the weights of the links between the household observing the events and the households where these events took place. Being the weights the inverse of the distance between the households' houses, the assumption is that the closer is the observed event the greater will be the increases in the perceived risk posed by the virus. On the other hand, when no event is observed, the probabilities are slowly reduced towards their 'rational' level, p_i, represented by the public information probability, through the equation:

$$p_{t+1} = p_t - \theta(p_t - p_i) \quad (5)$$

where θ is the rate at which the negative experience is forgotten by the agent.

The social influence. We assume that, for each venue with n attendants, the final isolation rate of an household, ${R_n}^f$, is the weighted average between its emotionally-biased-public-information isolation rate, ${R_n}^h$ and the weighted mean of the N neighbors' isolation rates, with the weights being those of the links between the household and its neighbors. Formally:

$$ {R_n}^f = \phi {R_n}^h + (1 - \phi) \sum_{i=1}^{N} w_i {R_n}^i \quad (6)$$

where ϕ is a parameter between 0 and 1 and the weights w_i are normalized so that their sum equals 1.

The virus course

Once an agent has become exposed, it is assigned one *infection course* over four possible courses: asymptomatic; symptomatic not hospitalized; hospitalized not in intensive care; in intensive care. In accordance with empirical studies, we assume that the probability of developing more serious conditions grows with age, decreases with social status and is higher for males than for females (Abate et al., 2020; Brazeau et al., 2020; Guilmoto, 2020; Yanez et al., 2020; England, 2020; Ferguson et al., 2020; Verity et al., 2020).

Upon exposure, the agent is also assigned an *incubation period*, which, in line with empirical observations (Lauer et al., 2020; McAloon et al., 2020), is drawn from a log-normal distribution with mean of about 5 days, and a *recovery period*, whose length depends on the severity level assigned to the agent, in order to reproduce a log-normal distribution of the recovery period with mean of about 12 days at the population level. The exposed agent is also assigned a *viral load*, which is drawn from a standard uniform distribution. The agent's viral load determines its contagiousness.

After the incubation period, the agent starts to develop symptoms (if not asymptomatic) and, in line with empirical observations (He et al., 2020), we assume that the exposed agent becomes infectious 2 days before the emergence of symptoms. The symptoms' severity of not-hospitalized symptomatic agents are differentiated by assigning them a *symptoms severity index*, between 0 and 1 exclusive, with the probability of the agent being assigned a higher value increasing with its viral load and its age, decreasing with its income quintile and being higher for males than for females. The closer to 1 the symptoms severity index, the more severe are the symptoms, the greater is the reduction of the agent's mobility. Therefore, the agent's mobility is given by the minimum between the illness-affected mobility and the mobility resulting from the behaviourally determined isolation rate of its household. Finally, the agent's symptoms severity index determine the probability that the agent will take a test.

After the recovery period, a share of agents will die, with a probability that also depends on age, social class and gender. All other agents recover and we assume that they are immune to Covid-19 thereafter.

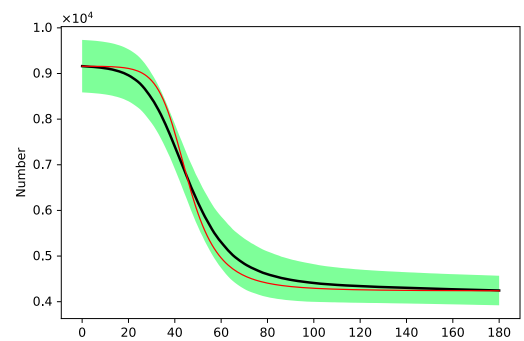

Figure 1: Susceptible agents.

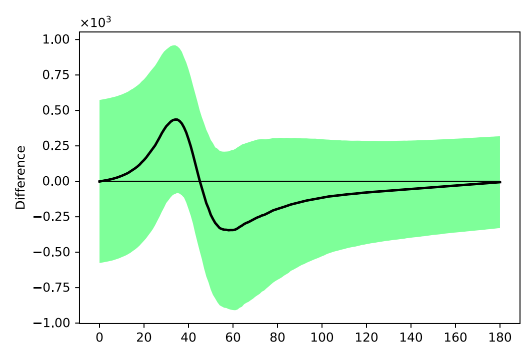

Figure 2: Susceptibles SEIR-simulation difference.

Simulation results

In this section, we present the results of a benchmark simulation repeated 20 times (the black lines representing the average and the green bands around it, the standard deviations). In Figure 1 we can observe the dynamics of the susceptible agents, compared to the typical logistic curve of a baseline $SEIR$ model with herd immunity. We can see that there is an initial period when the spread is faster than the baseline, then it slows to a lower rate than the baseline.

These dynamics are highlighted in Figure 2, showing the *difference* between the two curves of Figure 1. Here we can see that the spread in the ABM is faster than the baseline SEIR model (red line) up to around Day 40 and is lower thereafter, until around Day 60, when the SEIR curve becomes flatter than the simulation curve.

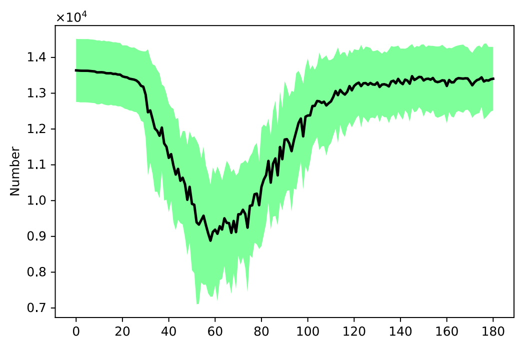

Figure 3: Total venue attendance.

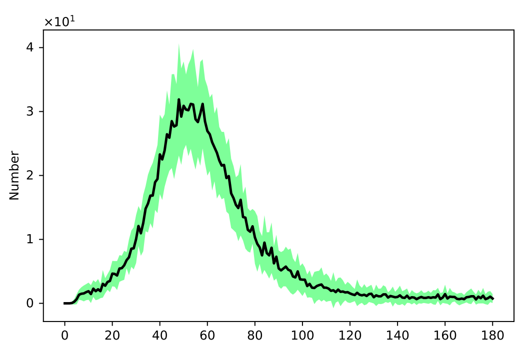

Figure 4: Number of new cases.

Figure 3 shows the total visits to venues, going from a maximum of around 1350 at the beginning of the simulation to a minimum of approximately 9000 at Day 60. We can see that there is a clear correspondence between the dynamics of venue attendance and the speed at which the virus spreads, represented by the number of new cases, shown in Figure 4.

In the next two figures, we show the number of hospitalizations. While Figure 5 shows the total number of people

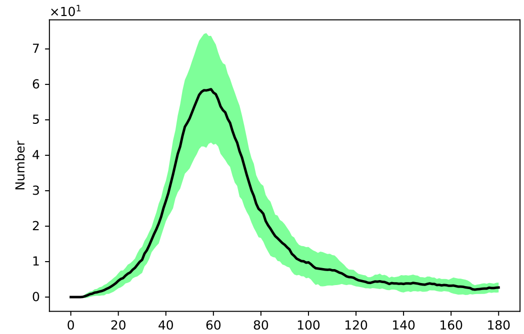

Figure 5: Hospitalized agents.

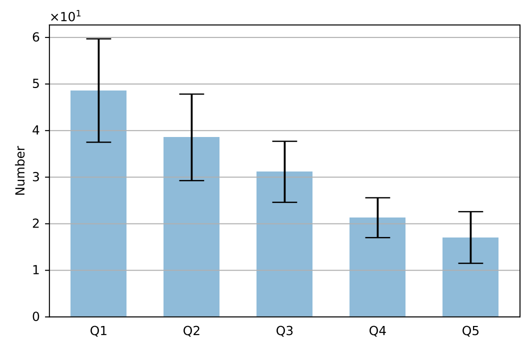

Figure 6: Hospitalizations by social class.

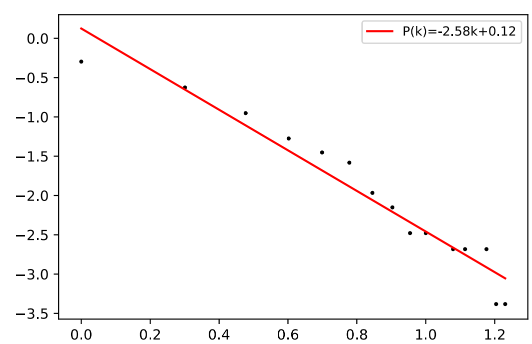

Figure 7: Degree distribution of infections' network.

hospitalized, Figure 6 shows that the effects of the pandemic are not equally distributed across income quintiles, with a clear negative social gradient: the highest quintile has, on average, little more than a third of the number of hospitalized agents than the lowest quintile.

Figure 7 shows the out-degree distribution of the infection network, where the degree represents the number of infected individuals by each contagious agent. We can see that the relationship between the frequency of degrees and the degrees on the log-log scale is approximately linear, (the R^2 being -0.97), with a scaling parameter of around 2.5, a typical value of scale-free networks. The fact that the model reproduces a scale-free infections network is rather surprising if we think that from a behavioural point of view all the agents have a similar behaviour (apart from 'linear' differences related to their age and socioeconomic status).

Conclusions

The simulations showed that the structure of the interactions and the agents' behavioral adaptations to the pandemic have effects that change the typical logistic curve which describes the virus spread in the traditional SEIR models. In this paper, we introduced a behavioural model through which the households decrease their social interaction as a response to the pandemic, a first step towards the development of more policy-relevant agent-based models of pandemics. We propose that models like this can provide a platform for designing and evaluating more effective public health restrictions and messaging during a pandemic. In future work, we will further explore the behaviour of the simulation with detailed sensitivity and uncertainty analyses, and investigate policy scenarios and their impact on behaviour, including lockdowns, social distancing and so on.

Acknowledgements

Umberto Gostoli and Eric Silverman are part of the Complexity in Health Improvement Programme supported by the Medical Research Council (MC_UU_00022/1) and the Chief Scientist Office (SPHSU16). This work was also supported by UK Prevention Research Partnership MR/S037594/1, which is funded by the the UK Research Councils, leading health charities, devolved administrations and the Department of Health and Social Care.

References

Abate, S. M., Ahmed Ali, S., Mantfardo, B., and Basu, B. (2020). Rate of intensive care unit admission and outcomes among patients with coronavirus: A systematic review and meta-analysis. *PloS one*, 15(7):e0235653.

Brazeau, N., Verity, R., Jenks, S., Fu, H., Whittaker, C., Winskill, P., Dorigatti, I., Walker, P., Riley, S., Schnekenberg, R. P., et al. (2020). Report 34: Covid-19 infection fatality ratio: estimates from seroprevalence.

England, P. H. (2020). Disparities in the risk and outcomes of covid-19.

Epstein, J. M. (2014). *Agent_Zero: toward neurocognitive foundations for generative social science*, volume 25. Princeton University Press.

Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., et al. (2020). Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. *Imperial College London*, 10(77482):491–497.

Green, L., Myerson, J., and Ostaszewski, P. (1999). Amount of reward has opposite effects on the discounting of delayed and probabilistic outcomes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2):418.

Guilmoto, C. Z. (2020). Covid-19 death rates by age and sex and the resulting mortality vulnerability of countries and regions in the world. *MedRxiv*.

He, X., Lau, E. H., Wu, P., Deng, X., Wang, J., Hao, X., Lau, Y. C., Wong, J. Y., Guan, Y., Tan, X., et al. (2020). Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675.

Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A. S., Reich, N. G., and Lessler, J. (2020). The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9):577–582.

McAloon, C., Collins, Á., Hunt, K., Barber, A., Byrne, A. W., Butler, F., Casey, M., Griffin, J., Lane, E., McEvoy, D., et al. (2020). Incubation period of covid-19: a rapid systematic review and meta-analysis of observational research. *BMJ open*, 10(8):e039652.

Prem, K., Cook, A. R., and Jit, M. (2017). Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS computational biology*, 13(9):e1005697.

Pullano, G., Valdano, E., Scarpa, N., Rubrichi, S., and Colizza, V. (2020). Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the covid-19 epidemic in france under lockdown: a population-based study. *The Lancet Digital Health*, 2(12):e638–e649.

Rescorla, R. A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, pages 64–99.

Silva, P. C., Batista, P. V., Lima, H. S., Alves, M. A., Guimarães, F. G., and Silva, R. C. (2020). Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons Fractals*, 139:110088.

Smith, L. E., Amlt, R., Lambert, H., Oliver, I., Robin, C., Yardley, L., and Rubin, G. J. (2020). Factors associated with adherence to self-isolation and lockdown measures in the uk: a cross-sectional survey. *Public Health*, 187:41–52.

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Cuomo-Dannenburg, G., Thompson, H., Walker, P. G., Fu, H., et al. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*, 20(6):669–677.

Wilder, B., Charpignon, M., Killian, J. A., Ou, H.-C., Mate, A., Jabbari, S., Perrault, A., Desai, A., Tambe, M., and Majumder, M. S. (2020). The role of age distribution and family structure on covid-19 dynamics: A preliminary modeling assessment for hubei and lombardy. In *SSRN*.

Yanez, N. D., Weiss, N. S., Romand, J.-A., and Treggiari, M. M. (2020). Covid-19 mortality risk for older men and women. *BMC Public Health*, 20(1):1–7.