

What causes second waves?

Karl J. Friston¹ and Alexander J. Billig²

¹*The Wellcome Centre for Human Neuroimaging, University College London, k.friston@ucl.ac.uk*

²*UCL Ear Institute, University College London, a.billig@ucl.ac.uk*

Summary

This report uses dynamic causal modelling of daily cases and deaths to distinguish between *secondary* and *second* waves, in terms of the underlying mechanisms and associated mortality due to SARS-CoV-2. Here, secondary waves are taken to reflect a spread of the virus through the population, reaching previously unexposed communities after the first wave has peaked. Conversely, second waves refer to the reinfection of a previously exposed population, due to a loss of effective (population) immunity. Our quantitative modelling suggests that the fatality rates that accompany both secondary and second waves are substantially less than the first wave—and an order of magnitude less than predictions based upon prevalent epidemiological models that ignore heterogeneity in viral exposure, susceptibility and transmission.

Introduction

There has been recent concern about what increases in new cases across Europe portend in terms of a second wave of SARS-CoV-2 infections. This brief report tries to clarify the mechanisms that might underwrite second waves and provide a quantitative estimate of their fatality. Although new cases have been increasing over the past weeks in several countries (at the time of writing: 3-Aug-20), hospital admissions and fatality rates have continued to decline^a. This is a remarkable observation that has a number of potential explanations. It could reflect an increase in the rate of testing—or its selectivity for infected cases—with no underlying increase in the prevalence of infection. Alternatively, it could reflect an increase in prevalence due to a relaxation of lockdown and travel restrictions; particularly among younger cohorts who have a lower infection fatality ratio¹. Another important distinction is between the spread of the epidemic into previously unexposed communities versus reinfection or resurgence in communities who have already been infected but have lost effective immunity (effective immunity here can be read in terms of host factors such as a decline in antibody levels^{2,3} or dilution of population immunity through population fluxes). In what follows, we operationalise this distinction by distinguishing between *secondary waves* involving a surge of infections—when the virus spreads to new communities—and *second waves* proper that represent a resurgence in a previously infected community. See Figure 1 for a famous example of second[ary] waves early last century.

We used dynamic causal modelling to disambiguate between these explanations using empirical timeseries of daily cases and deaths in several (largely European) countries. The dynamic causal model was equipped with heterogeneity of exposure, in that it distinguishes between people who constitute an exposed population and people who have yet to be exposed. This allows one to model the spread of the virus through the total population (e.g., country) in terms of movement from a *nonexposed* compartment

^a Note that hospital admissions and fatality rates do not cover all morbidity associated with COVID-19.

to an *exposed* compartment. Crucially, this is a nonlinear process that depends upon the prevalence of infection in the exposed population and the degree of mixing of the exposed and unexposed populations due to travel (e.g., commuting, tourism, and international flights). This means that the spread of the virus from the *exposed* to *unexposed* populations depends upon travel restrictions in exactly the same way that the spread of the virus within the *exposed* population depends upon lockdown measures.

Effectively, this dynamic causal model is a generalisation of previous models in which the effective population size *varies over time*, to model the spread of the virus from an initial to a final population size. The initial and final size are estimated proportions of the total (census) population size. This generalisation enables one to capture the epidemiological dynamics within an exposed population that grows in a self-assembling, autopoietic fashion, as it encroaches upon the nonexposed population. Furthermore, it allows one to model slow changes in model parameters as the effective population grows^b.

In dynamic causal modelling, lockdown and travel restrictions are treated as part of the epidemiological process. In other words, they become functions of various latent states, such as the prevalence of infection (as reflected in daily reports of new cases). Lockdown and travel restriction refer to behaviours rather than to regulations or guidance. The agenda is therefore to characterise how lockdown and travel restrictions are caused by viral transmission—and how viral transmission is caused by population behaviour. Once this circular causality has been parameterised, one can make predictions about what will happen in the future, using the best parameter estimates based upon past observations.

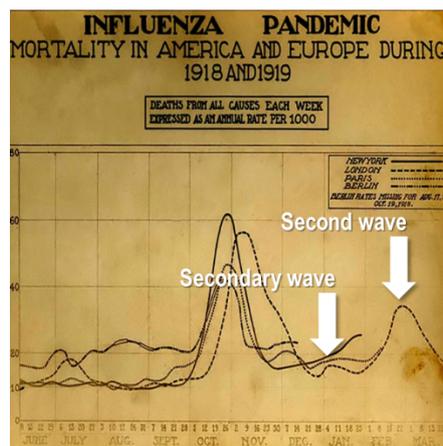


Figure 1: historical graphical data detailing mortality in America and Europe during the 1918/1919 Spanish flu epidemic. The fatality rates in London (dashed line) are highlighted with two arrows depicting a putative secondary and second wave, respectively. In this example, the second wave is about half the size of the first wave and ensues with the latency of about six months

^b In this model, the number of contacts in a high-risk setting was allowed to change with the size of the exposed or effective population to model spread from regions with high population density (e.g., megacities) to regions with a low population density (e.g., rural communities). In addition, the probability of dying given severe symptoms (c.f., the infection fatality ratio) was allowed to change, to model improvements in therapeutic management and any changing demographics of the exposed population.

Methods

We compared dynamic causal models in which *lockdown* was modelled in terms of the probability of leaving a low contact-risk location (e.g., home) for a high contact-risk location (e.g., work or the pub)⁴. Similarly, we modelled *containment* in terms of the spread of the virus through the total (census) population as the probability of leaving the *unexposed* population to enter the *exposed* population. These probabilities were (soft) threshold functions of the prevalence of infection. Crucially, the thresholds were allowed to differ for lockdown and containment (i.e., travel restrictions), respectively.

This threshold differential leads to a progression of the outbreak that has a fluctuating nature. Intuitively, as infection levels rise in an epicentre, travel restrictions contain the regional outbreak until a prevalence is reached that induces lockdown. At this point, the infection peaks, producing a first wave. As the prevalence of the virus declines, the exposed population unlocks, followed by a relaxation of travel restrictions. At this point, the infection ceases to be contained and spreads to other communities—and the process starts again. This leads to a fluctuating spread of the virus throughout the total population with secondary waves as new communities encounter the virus. When modelled, this process can repeat for several cycles until the virus has reached every corner of the population that it can access. Please see Figure 2 for a schematic describing this kind of model.

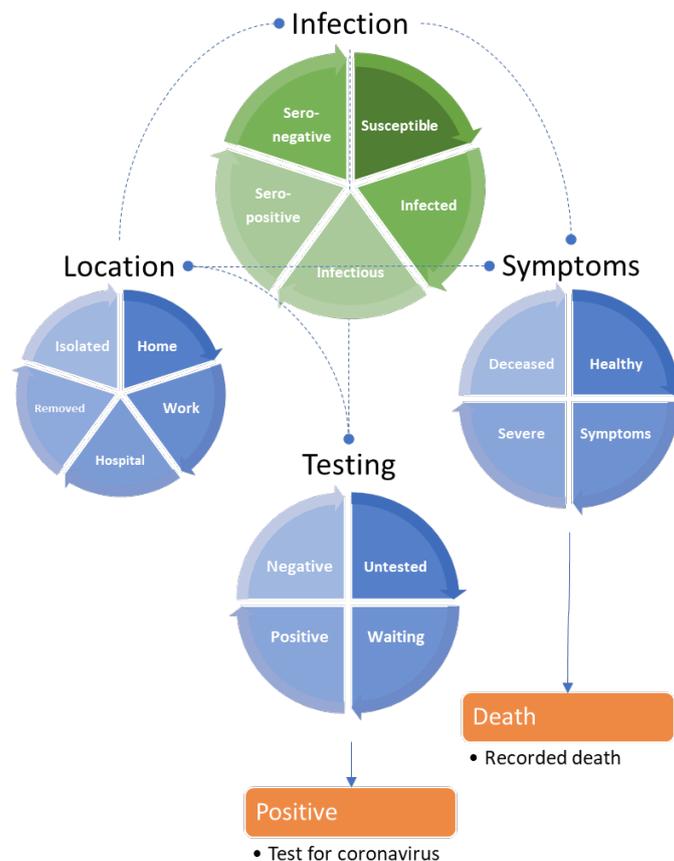


Figure 2: LIST model. This schematic summarises a LIST (*location, infection, symptom, and testing*) model used for the quantitative modelling. This model is formally equivalent to the model in ⁵. It includes a state (*isolated*) to model people who are shielded or self-isolating because they think they may be infectious. It also includes a (*seronegative*) state to model individuals with pre-existing immunity, e.g., via cross-reactivity ^{3,6} or other protective host factors ^{7,8}. This absorbing state plays the role of the *recovered* state of SEIR models—once entered, people stay in this state for the duration of the outbreak (e.g., long-lasting T-cell mediated immunity). The *removed* location state contains people who are not currently exposed to the epidemic. They slowly move to the *home* state in proportion to the prevalence of infection in the exposed population (i.e., *home* and *work*) and a containment (threshold) function of prevalence. The discs represent the four factors of the model, and the segments correspond to their states (i.e., compartments). The green disc is the closest to a conventional (i.e., SEIR) model that is augmented within three other factors. The states within any factor are mutually exclusive. In other words, every individual has to be in one state associated with each of four factors. The orange boxes represent the observable outputs generated by this model, in this instance, daily reports of positive tests and deaths. The rate of transition between states—or the dwell time within any state—rests upon the model parameters that, in many instances, can be specified with fairly precise prior densities—see Software and Data Note.

To build such a model, it is necessary to specify the precise functional form, not only of the way in which lockdown and containment depends upon the prevalence of infection, but how antigen testing generates new case data. This requires a careful treatment because the total number of tests is seldom available. This means that we have to model how new cases are generated without knowing the total number of tests or how selective these tests are for people who are infected. We tested a series of models in which the probability of being tested, when exposed, was a function of various factors (e.g., the proportion of people with symptoms, the proportion of the population exposed, seroprevalence, *etc.*). These models were fit to the eight countries that showed the most pronounced first wave using data from Johns Hopkins University (see Software and Data Note). The ensuing models were then optimised using Bayesian model reduction to remove redundant parameters: i.e., to identify the models with the greatest evidence that provide an accurate account of the data in the simplest way possible ⁹. Bayesian model reduction (a.k.a. structure learning) and Bayesian parameter averaging are completely automated (end-to-end) procedures, used routinely in dynamic causal modelling ^{10,11}.

Results

The resulting model had 34 parameters, which are listed with their prior means and covariances in `spm_SARS_priors.m`^c See Figure 3 for the resulting data fits and accompanying predictions. The Bayesian parameter averages over the optimised models were then used as empirical priors for fitting data from the United Kingdom.

^c <https://www.fil.ion.ucl.ac.uk/spm/covid-19/>

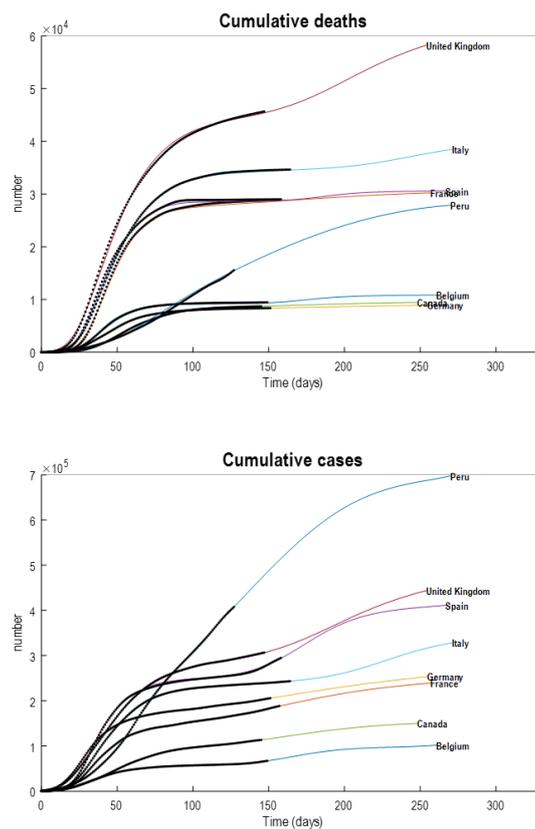


Figure 3: this figure illustrates the accuracy of model inversion by plotting the empirical data for cumulative deaths (upper panel) and cumulative new cases (lower panel) for several countries. The trajectories have been shifted in time such that zero weeks corresponds to the time point at which the prevalence of the infection was first estimated to be 0.1%. Note that these graphics include both empirical data (dots) and predictions (coloured lines) based upon the parameter estimates of the dynamic causal modelling (credible intervals have been omitted for clarity).

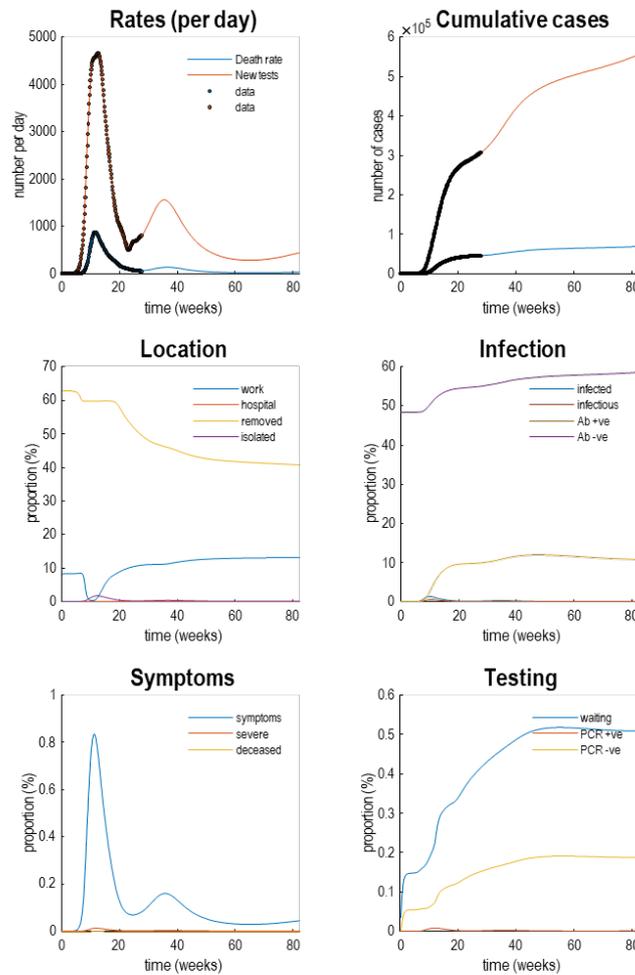


Figure 4: Latent causes. This figure illustrates posterior predictions of the most likely latent states for the United Kingdom with zero weeks corresponding to the time point at which the prevalence of infection was first estimated to be 0.1%. Proportions are given with respect to the total (census) population. Here, the outcomes in the upper two panels (dots) are supplemented with the underlying latent causes or expected states in the lower four panels (the first state in each factor has been omitted for clarity: i.e., *home*, *susceptible*, *healthy*, and *untested*). These latent or expected states generate the observable outcomes in the upper two panels. The solid lines are colour-coded and correspond to the states of the four factors in Figure 2. For example, under the *location* factor, the probability of being found at *work* (a high-risk setting—blue line) rises slightly as the virus spreads and then declines during lockdown. At this time, the probability of isolating oneself rises to about 2% during the peak of the pandemic. After about 20 weeks, containment is relaxed, and the unexposed populations becomes exposed with a fall in the proportion of *removed* people. At this time, the number of new cases starts to rise again with a slight increase in fatality rates, slowing down the unlocking process (reflected in the blue location line). Note that seroprevalence (yellow line in the *infection* panel) saturates at about 16 weeks (to roughly 10%) and is then further inflated by a few percent with a secondary wave. The seronegative state (*Ab -ve*: purple infection line) scores those people with pre-existing or humoral immunity (e.g. T-cell memory, induced by cross immunity with other coronaviruses or exposure to SARS-CoV-2) that does not entail seroconversion to *Ab +ve*. Please see Software and Data Note for details concerning the model inversion and data used to prepare this figure.

Figure 4 shows the data and most likely latent states for the United Kingdom. The equivalent posterior predictions for the reproduction ratio and fatality rates are shown as blue lines and Bayesian credible

intervals in Figure 5. Note that the secondary wave of new cases is not accompanied by a subsequent increase in fatality rates to the same extent as in the first wave. Please see figure legend for details.

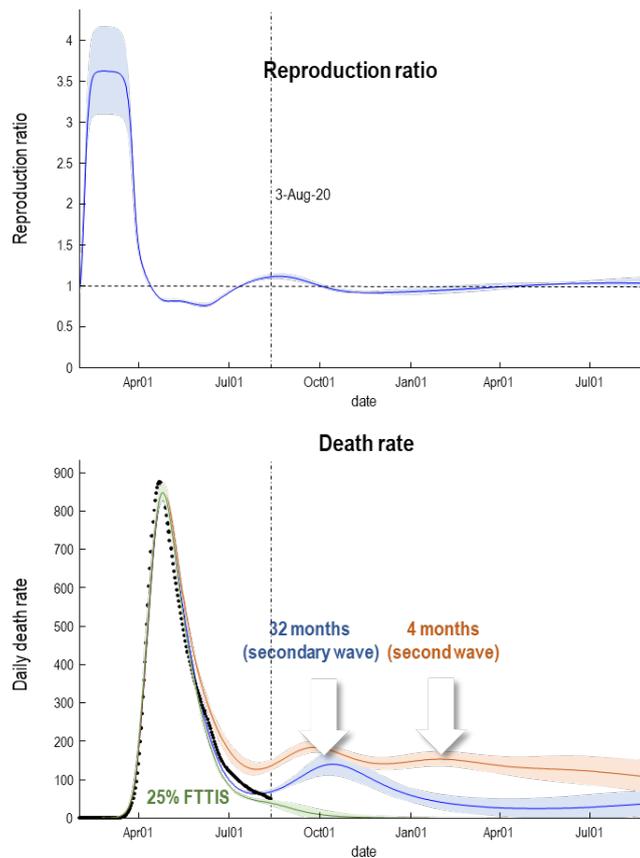


Figure 5: This figure reports the expected reproduction ratio (upper panel) death rates (lower panel) as a function of time for the United Kingdom, over an 18-month period. The three trajectories (lines) and accompanying 90% Bayesian credible intervals (shaded areas) correspond to posterior predictions. Note that the reproduction ratio is estimated to be about one at the present time but is predicted to fall to below one by October. The lower panel shows the equivalent fatality rates under three scenarios: a (prior) loss of immunity over 32 (blue) and four (orange) months, respectively and under an enhanced testing and tracing protocol (green) for enduring immunity over 32 months. The green line reproduces the predictions based upon the posterior parameter estimates for the United Kingdom but increasing the efficacy of find, test, trace, isolate and support to 25% (from its posterior estimate of about 1%). It can be seen that even with this relatively low efficacy, elimination is possible by November, with convergence to zero fatality rates.

The above modelling assumed a period of enduring (effective) immunity with the prior expectation of 32 months^d, specific to seropositive cases who seroconvert following infection. If we now reduce this period to 4 months but retain the remaining parameter estimates, we can generate second waves of the sort previously considered, in the absence of secondary waves¹²⁻¹⁴. The orange line and credible intervals in Figure 5 report the equivalent predictions under a relatively rapid loss of immunity—that renders people who have seroconverted susceptible to re-infection. The result is a protracted second wave, accounting for several thousand extra fatalities.

^d Strictly speaking, the two months is a time constant or inverse rate constant.

Notably, this kind of second wave arises slightly earlier than can be explained just in terms of loss of immunity¹⁴ (models without a slowly increasing effective population, show a second wave peaking early in January). This speaks to the interaction between viral spread and loss of immunity, in conspiring to accelerate the emergence of a second wave.

Another notable aspect of these posterior predictions is that, quantitatively speaking, the second wave entails a smaller mortality than the first. This should be contrasted with predictions based upon agent-based simulation models that ignore heterogeneity. Most of these models suggest that the number of people dying in the second wave will exceed the fatalities of the first^{15,16}. The Academy of Medical Science in their 14-July-20 reportⁱ ‘Preparing for a challenging winter 2020/21’ suggest a peak in hospital admissions and deaths in January/February 2021 with estimates of 119,900 (95% credible interval 24,500-251,000) hospital deaths between September 2020 and June 2021—double the number that occurred during the first wave in the spring of 2020. This is an almost impossible scenario under models that consider heterogeneity of variation in the exposure to the virus, the susceptibility to infection and the capacity to transmit the virus when infected^e. Having said this, the particular model of heterogeneity adopted above is somewhat minimal, in the sense that it does not model variations over stratified age ranges. Furthermore, we have not considered seasonal variations in transmission or interaction with things like seasonal influenza^{12,19}.

The third and final prediction (green) in Figure 5 is the equivalent posterior prediction if an effective find, test, trace, isolate, and support (FTTIS) scheme is put in place from this point onwards in the United Kingdom. Here, an effective FTTIS scheme was taken to render the probability of somebody self-isolating—when infected but before they became infectious—25%. In principle, this level of efficacy is practical²⁰⁻²² and, according to the above quantitative modelling, capable of suppressing community transmission by autumn.

Conclusion

The current increases in daily cases in many European countries probably reflects a relaxation of travel restrictions (i.e., *containment* strategies) both within and between countries and the denouement of viral spread into hitherto unaffected communities. In the context of enduring immunity, it is likely that these secondary fluctuations will disappear over the next month or two—and will not be accompanied by a secondary increase in mortality rates. However, if people who have been infected lose their immunity in the short term (e.g., over four months) then these secondary fluctuations will merge into a second wave in November, with the potential loss of several thousand (but not tens of thousands) of lives in the United Kingdom. This loss could, in principle, be mitigated by testing and tracing, if implemented efficiently at the present time (August 2020). As can be seen from the quantitative modelling in Figure 5, there is every possibility of suppressing community transmission (i.e., attaining Zero-COVID) with suitably enhanced testing and tracing by the end of the calendar year.

^e The discrepancy between conventional (agent-based) models and dynamic causal models is possibly a reflection of the fact that conventional models have to commit to rather simplistic assumptions, because it is difficult to compare competing models – and therefore improve the model with accumulating data or hypotheses. Stochastic transmission models are notoriously difficult to evaluate in terms of their evidence¹⁷. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International journal of epidemiology* 2006; **35**(3): 765-75, 18. Chatzilena A, van Leeuwen E, Ratmann O, Baguelin M, Demiris N. Contemporary statistical inference for infectious disease models using Stan. *Epidemics* 2019; **29**: 100367. In contrast, the variational approaches used in dynamic causal modelling furnish a variational bound on model evidence that allows competing models to be assessed quickly and efficiently.

Acknowledgements

The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome Trust [203147/Z/16/Z].

Author declarations

The authors declare no conflicts of interest.

Software and data note

The analyses in this article can be reproduced using annotated (MATLAB) code available as part of the open source academic software SPM (<https://www.fil.ion.ucl.ac.uk/spm/>), released under the terms of the GNU General Public License version 2 or later. The routines are called by a demonstration script: DEM_SARS.m. Please visit <https://www.fil.ion.ucl.ac.uk/spm/covid-19/>. The data used in this article are available for academic research purposes from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, hosted on GitHub at <https://github.com/CSSEGISandData/COVID-19>.

References

1. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious diseases* 2020.
2. Stringhini S, Wisniak A, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *The Lancet* 2020.
3. Ng K, Faulkner N, Cornish G, et al. Pre-existing and de novo humoral immunity to SARS-CoV-2 in humans. *bioRxiv* 2020: 2020.05.14.095414.
4. Friston K, Parr T, Zeidman P, et al. Dynamic causal modelling of COVID-19 [version 1; peer review: awaiting peer review]. *Wellcome Open Research* 2020; **5**(89).
5. Friston KJ, Parr T, Zeidman P, et al. Tracking and tracing in the UK: a dynamic causal modelling study. *arXiv e-prints* 2020: arXiv:2005.07994.
6. Grifoni A, Weiskopf D, Ramirez SI, et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* 2020.
7. Zheng M, Gao Y, Wang G, et al. Functional exhaustion of antiviral lymphocytes in COVID-19 patients. *Cellular & Molecular Immunology* 2020; **17**(5): 533-5.
8. Bunyavanich S, Do A, Vicencio A. Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults. *JAMA* 2020.
9. Friston K, Parr T, Zeidman P. Bayesian model reduction. *arXiv preprint arXiv:180507092* 2018.
10. Litvak V, Garrido M, Zeidman P, Friston K. Empirical Bayes for Group (DCM) Studies: A Reproducibility Study. *Front Hum Neurosci* 2015; **9**: 670.
11. Friston KJ, Litvak V, Oswal A, et al. Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage* 2016; **128**: 413-31.

12. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 2020: eabb5793.
13. Friston K, Parr T, Zeidman P, et al. Second waves, social distancing, and the spread of COVID-19 across America [version 1; peer review: awaiting peer review]. *Wellcome Open Research* 2020; **5**(103).
14. Friston KJ, Parr T, Zeidman P, et al. Effective immunity and second waves: a dynamic causal modelling study. 2020.
15. Davies NG, Kucharski AJ, Eggo RM, et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *The Lancet Public Health* 2020; **5**(7): e375-e85.
16. Okell LC, Verity R, Watson OJ, et al. Have deaths from COVID-19 in Europe plateaued due to herd immunity? *The Lancet*.
17. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *International journal of epidemiology* 2006; **35**(3): 765-75.
18. Chatzilena A, van Leeuwen E, Ratmann O, Baguelin M, Demiris N. Contemporary statistical inference for infectious disease models using Stan. *Epidemics* 2019; **29**: 100367.
19. Lourenco J, Pinotti F, Thompson C, Gupta S. The impact of host resistance on cumulative mortality and the threshold of herd immunity for SARS-CoV-2. *medRxiv* 2020: 2020.07.15.20154294.
20. Kretzschmar M, Rozhnova G, van Boven M. Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. *Available at SSRN* 3562458 2020.
21. Keeling MJ, Hollingsworth TD, Read JM. The Efficacy of Contact Tracing for the Containment of the 2019 Novel Coronavirus (COVID-19). *medRxiv* 2020.
22. Hellewell J, Abbott S, Gimma A, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health* 2020; **8**(4): e488-e96.

ⁱ Government Publications: Research and analysis: COVID-19: Preparing for a challenging winter 2020/21, 7 July 2020 (Paper prepared by the Academy of Medical Sciences) <https://acmedsci.ac.uk/file-download/51353957>