Modeling the epidemic of the Covid-19 omicron variant in England

Philip Thomas*

Safety Systems Research Centre, South West Nuclear Hub, Faculty of Engineering, University of Bristol, Queen's Building, University Walk, Bristol BS8 1TR, UK

The model of the Covid-19 epidemic in England has been extended to allow for two co-existing variants, delta and the new omicron strain. A simple model is proposed to assess the likely beneficial effect of the government's booster campaign of third vaccine doses. Details of the new modeling required are given. The model results are compared with incoming data from both the Health Security Agency and the Office of National Statistics, and found to hold up well. Peaks are predicted for active cases in England, 3.7 million in the base case, 3900 hospitalizations per day, and 500 deaths per day. The latter two figures are below what England saw in January 2021. A sensitivity study raises these numbers by between 50 and 70%, to 5.6 million, 6500 and 850, respectively. However, South Africa, which faced the omicron strain first, has reported that hospital stays with omicron are half what they experienced with other strains. Bed occupancy would stay within tolerable limits if this behaviour translated to England, even with the rate of hospitalization expected in the sensitivity study. A postscript has been added to cover new data up to 4 March 2022, which provides validation for the model used. A discussion is given of the poor performance of the government's official modeling groups and the damage this has caused to the reputation of epidemic modeling as a field. This is obviously unfortunate when it is clear that the future course of the Covid-19 epidemic could have been modeled accurately, as explained in this paper, and there is no reason why an epidemic of a new disease should not be modeled accurately in the future. But epidemic models, if they are to be useful in controlling an epidemic like Covid-19, need to be updated daily to allow decision makers to have the earliest possible notification of the state of the epidemic that they are seeking to control. A further eight lessons are listed.

Keywords: delta variant, lessons learned, pandemic, validation

* E-mail: philip.thomas@bristol.ac.uk; research website: www.jvalue.co.uk

Nanotechnology Perceptions **18** (2022) 89–115
Received 6 January 2022; *accepted* 20 February 2022

doi: 10.4024/N02TH22A.ntp.18.02

1. Introduction

The coronavirus disease 2019, Covid-19, which was first reported in Wuhan, China in December 2019, quickly became a worldwide pandemic. To date there have been more than 287 million reported cases and 5.45 million deaths. The omicron variant was first documented on 9 November 2021 in the city of Tshwane, South Africa, an urban area centred on Pretoria with a population of 3.3 million. The new variant led to exponential increases in cases and a sharp rise in hospital admissions there, before reaching a peak in early December 2021, after which it began to decline.

The first cases of the omicron variant were detected in England on 27 November 2021. The variant spread very rapidly, so that omicron became the dominant strain by late December 2021, taking over from the delta variant, even though delta infections persisted in very large numbers.

South African reports suggested that the omicron variant was more infectious than other strains, but that its symptoms were significantly milder.[1] Those unlucky enough to need hospitalization with omicron spent only 4 days in hospital on average, compared with 9 days during previous waves of other variants.

While the Look-Ahead Predictor Corrector Coronavirus Filter (LAPCCF)[2] proved to be a good predictor of the spread of the delta variant in England, predictions after the middle of December needed to take proper account of the incursion of the new omicron variant, as it challenged the preexisting delta variant for dominance. The framework used previously was extended to allow for two variants, delta and omicron, progressing simultaneously.

2. Changes to the model

The bracketed superscript, [k], is introduced to denote the type of variant, with delta being denoted by $k = 1$ and omicron by $k = 2$. This new superscript adds to the previous superscript $j = 1, 2, ..., 8$ used to distinguish between age groups, and the subscript $i = 1, 2$ used to denote symptom severity.

The two variants, delta and omicron, progress through the population simultaneously but the spreads are not independent, because it is assumed that:

(i) vaccination will cause a fraction $\eta_{imm}^{[k]}$ of the people to become immune to variant k. Hence the fraction who become immune to delta will be $\eta_{imm}^{[1]}$, while the fraction who become immune to omicron will be $\eta_{imm}^{[2]}(t)$, where time t has been included to allow for the effect of third ("booster") vaccinations increasing the fraction immune to the omicron variant over time. A cross-variant coefficient, $\eta_{imm}^{[1][k]}(t)$, may be defined relative to the fraction becoming immune to delta infection ($k = 1$):

$$\eta_{imm}^{[1][k]}(t) = \frac{\eta_{imm}^{[k]}(t)}{\eta_{imm}^{[1]}} \qquad \text{for } k = 1, 2. \tag{1}$$

Clearly $\eta_{imm}^{[1][1]}(t) = \eta_{imm}^{[1]} / \eta_{imm}^{[1]} = 1$, while the lower effectiveness of the vaccines against the omicron variant may be characterized by a cross-variant immunity coefficient that is less than unity: $\eta_{imm}^{[1][2]}(t) = \eta_{imm}^{[2]}(t) / \eta_{imm}^{[1]} < 1$.

---

[1] F. Abdullah et al., Decreased severity of disease during the first global omicron variant covid-19 outbreak in a large hospital in Tshwane, South Africa. *Intl J. Infectious Diseases* **116** (2022) P38–P42.

[2] P. Thomas, Vaccine-mediated exit strategies from England's Covid-19 lockdown. *Nanotechnol. Perceptions* **17** (2021) 30–73.

(ii) recovery from an omicron infection brings about immunity to future infections, both omicron and delta, at least in the near future.

(iii) people who recover from a delta infection will be immune to a future delta infection, at least in the near future, but only a fraction, $\eta_{imm}^{[1][2]}(t)$, of those who recover from a delta infection will be immune to a future omicron infection. $\eta_{imm}^{[1][2]}(t)$ is taken to be the same as the cross-variant immunity coefficient for vaccination, given by eqn (1).

The effect of the government's booster campaign, which converts relatively low levels of omicron immunity due to vaccination into much higher levels of resistance, is then simulated by assuming that the cross-variant immunity coefficient $\eta_{imm}^{[1][2]}(t)$ will increase linearly from 60%, in late November 2021 to 90% by early January 2022. A sensitivity study examines the case where attaining the 90% figure for $\eta_{imm}^{[1][2]}(t)$ is delayed until the end of January 2022.

The simulation is effected by modifying the LAPCCF to contain:

(i) a set of modules representing the progress of the delta variant, which are unchanged from the original version described by Thomas (2021),[1] except that they now include cross-variant terms, plus:

(ii) a new set of mainly independent but conceptually similar modules to characterize the progress of the omicron variant.

The vaccination calculations are retained within the delta modules, and the flows of vaccinated people who are immune to delta are calculated as before (see eqn A.23 of ref. 2). Meanwhile the flows $q_i^{(j)[2]}(t)$ of newly vaccinated people in cohort *i* and age group *j* who develop immunity to omicron are given in terms of the flows of those developing immunity from delta, $q_i^{(j)[1]}(t)$, by:

$$q_i^{(j)[2]}(t) = \eta_{imm}^{[1][2]}(t)\, q_i^{(j)[1]}(t) \qquad \text{for } i = 1, 2;\ j = 1, 2, ..., 8. \tag{2}$$

Those classed as recovered include not only the people who have recovered from infection but also those who become immune through vaccination. Thus the rate of growth in the number immune to infection from the delta variant, where $k = 1$, now needs to contain an extra, final term, $\mathrm{d}n_{ri}^{(j)[2]}/\mathrm{d}t\,|_{pure}$, to represent the cross-variant effect of those recovering from omicron who are assumed to gain immunity against delta, also:

$$\frac{\mathrm{d}n_{ri}^{(j)[1]}}{\mathrm{d}t} = \left.\frac{\mathrm{d}n_{ri}^{(j)[1]}}{\mathrm{d}t}\right|_{pure} + q_i^{(j)[1]}(t) + \left.\frac{\mathrm{d}n_{ri}^{(j)[2]}}{\mathrm{d}t}\right|_{pure} \qquad i = 1, 2;\ j = 1, 2, ..., 8. \tag{3}$$

Here the pure recovery terms are given by:

$$\left.\frac{\mathrm{d}n_{ri}^{(j)[k]}}{\mathrm{d}t}\right|_{pure} = \frac{n_i^{(j)[k]}(t)}{\tau_{\mathrm{inf},i}} \qquad i = 1, 2;\ j = 1, 2, ..., 8. \tag{4}$$

For the omicron variant, where $k = 2$, the equations now take the form:

$$\frac{\mathrm{d}n_{ri}^{(j)[2]}}{\mathrm{d}t} = \left.\frac{\mathrm{d}n_{ri}^{(j)[2]}}{\mathrm{d}t}\right|_{pure} + q_i^{(j)[2]}(t) + \eta_{imm}^{[1][2]}(t)\left.\frac{\mathrm{d}n_{ri}^{(j)[1]}}{\mathrm{d}t}\right|_{pure} \qquad i = 1, 2;\ j = 1, 2, ..., 8 \tag{5}$$

where $q_i^{(j)[2]}(t)$ is as given in eqn (2) above. The final term in eqn (5) represents the cross-variant effect of a fraction $\eta_{imm}^{[1][2]}(t)$ of the delta recoveries who are becoming immune not only to delta but also to omicron. The assumption made, for simplicity, is that the fraction of delta

recoveries who are immune to omicron also will increase over time (see §3, assumption 2); the inaccuracy introduced will be small because (i) delta recoveries will soon be a great deal fewer than omicron recoveries as the omicron variant takes off; (ii) people recovering from delta will, in any case, soon receive a third vaccination.

The total number $N_i^{(j)}$ in age group j and symptom cohort i will be given as $\theta_i p^{(j)} N$, where N is the total number in the population $p^{(j)}$, $j = 1, 2, ..., 8$ is the fraction of people in each of the 8 age groups and θ_i, $i = 1, 2$ is the fraction in each of the symptom cohorts. The number of people with an active infection will be $\sum_{k=1}^{2} n_i^{(j)[k]}(t)$ for symptom cohort i and age group j.

Assuming no one becomes infected with either variant while they are actually combating an infection, the number of people susceptible to the kth variant may be found by subtraction:

$$n_{si}^{(j)[k]}(t) = \theta_i p^{(j)} N - n_{ri}^{(j)[k]}(t) - \sum_{k=1}^{2} n_i^{(j)[k]}(t) \qquad i = 1, 2;\ j = 1, 2, ..., 8;\ k = 1, 2. \tag{6}$$

3. Summary of assumptions

1. Those with T-cell immunity gained as a result of a prior coronavirus infection, estimated to be 12.9% of the population, will be immune to both delta and omicron variants.

2. The resistance of the population of England to infection with the omicron variant on date t_{00} = 21 November 2021, a week before the first omicron cases were detected, as a result of both vaccination and recovery from another Covid variant, is 60% of what it was against delta. However the booster campaign causes this figure to rise to about 90% by t_{ff} = 7 January 2022.

3. Those who contract the omicron variant and recover will have full resistance to infection from delta.[3]

4. The active infections predicted into the future by the LAPCCF model consist of both delta and omicron infections, both of which are calculated in the model.

5. Official figures for cases by date reported imply infections that took place 4 days before.

6. The number of active omicron infections is small compared with those from the delta variant up to the end of 7 December 2021.

7. The time between infection and transmission for the omicron variant is the same as found previously for the ancestral Wuhan variant, namely 8.82 days for symptom cohort 1, and 8.25 days for symptom cohort 2,[4] in both cases across all ages.

8. The LAPCCF results are matched to observed data over the interval 28 May–7 December 2021 by adjusting the Social Mixing Index for the delta variant and the average length of time between infection and transmission for delta to minimize the delta-variant objective function described next.

9. The delta-variant objective function consists of the addition of the two components:

(i) the sum of the squared errors between the daily cases, based on official data after correction to a standard number of daily tests, and the predicted daily corrected cases, between the observation dates 1 June 2021 and 11 December 2021, equivalent to actual dates 28 May

[3] S. Knapton, Catching omicron Covid variant may protect against delta. *Daily Telegraph* (28 December 2021).

[4] P. Thomas, The length and severity of the coronavirus recession estimated from the dynamics of relaxing lockdown. *Nanotechnol. Perceptions* **16** (2020) 100–129.

2021 and 7 December 2021; (ii) the weighted sum of the squared errors between the ONS figure for the number of active infections[5] and those predicted from 28 May 2021 up to 7 December 2021.

10. The total new daily cases predicted by the model constitute a best estimate of the numbers of new daily delta cases in future days.

11. Most cases are captured in the official statistics for daily new cases for the interval of observation dates 12–20 December 2021 (equivalent to actual dates of infections 8–16 December).

12. Restrictions in place between 8 and 16 December 2021 continue into the future, so that no new countermeasures more severe than "Plan B" are introduced. It is assumed that no new restrictions, short of severe lockdown, would have a significant effect, even if imposed, because the omicron variant is so infectious.

13. The number recovering from omicron over the interval 8–16 December 2021 will be small compared with the number becoming infected—reasonable when infections are rising very rapidly.

14. The omicron hospitalization rate is one third of the delta hospitalization rate.[6,7]

15. Those who die of Covid will have been hospitalized first.

16. 1 in 7.5 people who are hospitalized with omicron will die of Covid, the same figure as for delta (although new reports coming from South Africa suggest that this may be pessimistic).[1]

17. People will not become infected with either variant during the time that they are actually combating an infection caused by either the delta variant or the omicron variant. These will number $\sum_{k=1}^{2} n_i^{(j)[k]}(t)$ for symptom cohort i and age group j. Meanwhile the total number in the age group and symptom cohort, $N_i^{(j)}$, is, as explained in the second-last paragraph of §2, $\theta_i p^{(j)} N$. Hence the number of people susceptible to the kth variant may be found by subtraction, as given by eqn (6), repeated below:

$$n_{si}^{(j)[k]}(t) = \theta_i p^{(j)} N - n_{ri}^{(j)[k]}(t) - \sum_{k=1}^{2} n_i^{(j)[k]}(t) \quad i = 1, 2;\ j = 1, 2, ..., 8;\ k = 1, 2. \tag{6}$$

Assumptions 15, 16 and 17 taken together imply a death rate for omicron that is a third of that for delta.

4. Estimating the Social Mixing Index for the omicron variant

Benefit was taken from the fact that both the LAPCCF, without the extra omicron modules, and the original measurement model, the Predictor Corrector Coronavirus Filter (PCCF), had been matched, every day for six months up to the beginning of December 2021, to figures for cases by date reported daily by the Health Security Agency and to the numbers of active infections in

---

[5] Office of National Statistics (ONS), Coronavirus (COVID-19) Infection Survey headline results, UK (5 January 2022) https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/5january2022

[6] National Institute for Communicable Diseases, South Africa, Covid-19 hospital surveillance update: week 48, December 2021.

[7] A. Sheikh, S. Kerr, M. Woolhouse, J. McMenamin & C. Robertson, Severity of omicron variant of concern and vaccine effectiveness against symptomatic disease: national symptom cohort with nested test negative design study in Scotland (2021). *The Lancet Infectious Diseases* **22** (2022) 959–966**.**

England, supplied weekly in arrears by the Office of National Statistics. The LAPCCF had proved a good predictor of the course of the delta epidemic to at least 10 days into the future.

The assumption was made that the observed data corresponding to the interval 28 May–7 December 2021 were caused predominantly by delta infections, and the LAPCCF results were matched to these by adjusting the Social Mixing Index (SMI), $\Sigma^{[1]}$, for the delta variant and the average length of time between infection and transmission, $\tau_{\text{inf},i}^{[1]}$, $i = 1, 2$, again for the delta variant, with the aim of minimizing a delta-variant objective function, which consisted of the addition of two components:

(i) the sum of the squared errors between the daily cases, based on official data (after correction to a standard number of daily tests), and the predicted daily corrected cases, between the observation dates 1 June and 11 December 2021, equivalent to actual dates 28 May and 7 December 2021;

(ii) the weighted sum of the squared errors between the ONS figure for the number of active infections and those predicted from 28 May up to 7 December 2021.

The new daily cases predicted into the future by the model were then held to be best estimates of the numbers of new daily delta cases. The number of daily new omicron cases over the interval 8–16 December 2021 were then computed as the difference between the official figures for new cases per day (observed between 12 and 20 December, but considered to be valid 8–16 December, by a previous assumption), and the LAPCCF best estimates of daily new delta cases (uncorrected "cases by date reported") over the interval 8–16 December.

The estimated number of active omicron infections 8–16 December were then found by integrating the daily new omicron cases just found, from an initial condition for the number of active omicron infections at the end of 7 December 2021, using the subsidiary assumption that the number recovering from omicron over the interval would be small compared with the infection rate—reasonable when infections were rising very rapidly from a low base.

The initial condition for active omicron infections and the omicron SMI were first guessed and then varied so as to minimize a further objective function, which consisted of the sum of the squared errors between the numbers of omicron active infections calculated by the LAPCCF over the interval 8–16 December and the values estimated as described above.

The initial condition for active omicron infections in England on 7 December 2021 was found to be 35 200, and the SMI for the omicron variant was estimated to be 9.19, 23–53% higher than for the already very infectious delta strain, the SMI for which is estimated to lie in the range 6 to 7.5. It implies that the average person carrying an omicron infection in a population with no prior immunity would infect over nine people before recovering. Taken in conjunction with its ability to infect people with immunity to delta, this guaranteed that omicron would rapidly become the dominant strain in England, and indeed the world, despite the number of omicron active infections being only about a thirtieth of the figure for delta active infections on 7 December.

The initial condition of 35 200 active infections at the end of 7 December 2021 was distributed in proportion to the fraction θ_i of people in each symptom cohort, and the resulting numbers were inserted as initial conditions into the module covering the age group 25–49, which is the most populous and also the one where most omicron cases were observed to begin with.

The final match between derived and computed active infections in England over the interval 8–16 December 2021 is shown in Fig. 1.

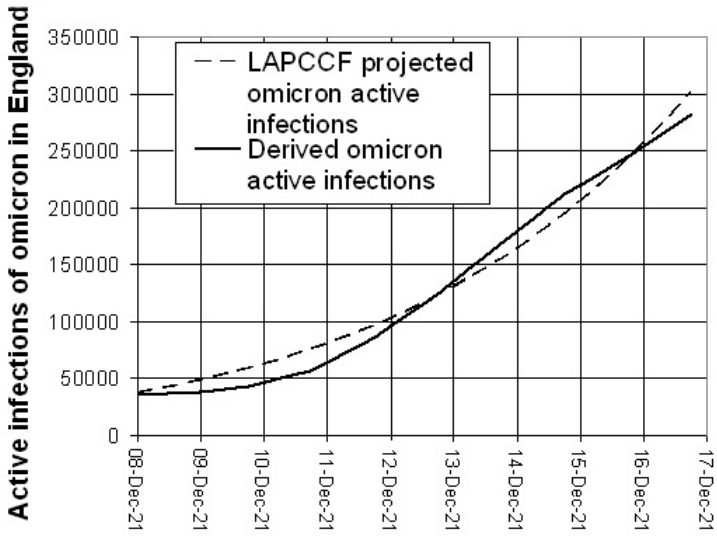

Figure 1. Matching estimated omicron active infections over the interval 8 December to 16 December 2021.

5. The number in the recovered category who are immune to omicron

Let t_{00} be a date just before the entry of the omicron variant into England. As of that date, the people in each age group and symptom cohort who were estimated to be immune to omicron were composed of:

(i) those with prior T-cell immunity to omicron, who numbered $0.129N_i^{(j)}$, where $N_i^{(j)} = \theta_1 N^{(j)}$ is the number in symptom cohort i and age group j;

(ii) those who are immune to omicron as a result of either vaccination or recovery from either delta or an earlier variant. This number may be calculated as

$$\eta_{imm}^{[1][2]}(t_{00})\left(n_{ri}^{(j)[1]}(t_{00}) - 0.129N_i^{(j)}\right) \tag{7}$$

where $n_{ri}^{(j)[1]}(t_{00})$ is the number immune to delta as a result of either vaccination or recovery.

Adding these two numbers together gives the initial condition for those in the recovered category for omicron, and hence immune to omicron as:

$$n_{ri}^{(j)[2]}(t_{00}) = 0.129N_i^{(j)} + \eta_{imm}^{[1][2]}(t_{00})\left(n_{ri}^{(j)[1]}(t_{00}) - 0.129N_i^{(j)}\right) \quad \text{for } i = 1, 2;\ j = 1, 2, ..., 8. \tag{8}$$

We may define the difference $\Delta n_{ri}^{(j)[2]}$ between the number immune to omicron and the corresponding number immune to delta at time t_{00} by:

$$\Delta n_{ri}^{(j)[2]} = n_{ri}^{(j)[1]}(t_{00}) - n_{ri}^{(j)[2]}(t_{00}) \quad \text{for } i = 1, 2;\ j = 1, 2, ..., 8. \tag{9}$$

In a similar fashion, let $Q_i^{(j)[1]}(t_{00})$ be the number rendered immune by vaccination to the delta variant at time t_{00}, given by integrating the flow of people newly immune to delta, $q_i^{(j)[1]}(t)$, from the time t_{v0} when the vaccination campaign began in December 2020:

$$Q_i^{(j)[1]}(t_{00}) = \int_{t=t_{v0}}^{t_{00}} q_i^{(j)[1]}(t)\,\mathrm{d}t \qquad \text{for } i = 1, 2;\ j = 1, 2, ..., 8. \tag{10}$$

The number of people who can be considered immune to the omicron variant as a result of this vaccination campaign at time t_{00} will then be:

$$Q_i^{(j)[2]}(t_{00}) = \eta_{imm}^{[1][2]}(t_{00}) Q_i^{(j)[1]}(t_{00}) \quad \text{for } i = 1, 2;\ j = 1, 2, ..., 8. \tag{11}$$

We may define the difference between these two numbers as $\Delta Q_i^{(j)[2]}$:

$$\Delta Q_i^{(j)[2]} = Q_i^{(j)[1]}(t_{00}) - Q_i^{(j)[2]}(t_{00}) \quad \text{for } i = 1, 2;\ j = 1, 2, ..., 8. \tag{12}$$

Finally, for completeness, the number of people who fall into the purely recovered category and are immune to omicron across both symptom cohorts will be:

$$n_r^{(j)[2]}(t_{00})\Big|_{pure} = \sum_{i=1}^{2} n_{ri}^{(j)[2]}(t_{00}) - 0.129N - \sum_{i=1}^{2} Q_i^{(j)[2]}(t_{00}) \qquad \text{for } j = 1, 2, ..., 8. \tag{13}$$

The difference, $\Delta n_r^{(j)[2]}\Big|_{pure}$, is then defined as:

$$\Delta n_r^{(j)[2]}\Big|_{pure} = n_r^{(j)[1]}(t_{00})\Big|_{pure} - n_r^{(j)[2]}(t_{00})\Big|_{pure}. \tag{14}$$

The booster campaign is assumed to restore the population resistance by adding to the time differentials of $n_{ri}^{(j)[2]}(t)$, $Q_i^{(j)[1]}(t_{00})$ and $n_r^{(j)[2]}(t_{00})\Big|_{pure}$ the respective terms:

$$\begin{cases} \dfrac{\eta_{imm}^{[1][2]}(t_{ff}) - \eta_{imm}^{[1][2]}(t_{00})}{1 - \eta_{imm}^{[1][2]}(t_{00})} \dfrac{\Delta n_{ri}^{(j)[2]}}{t_{ff} - t_{00}} & \text{for} \quad t_{00} < t \le t_{ff} \\ 0 & \text{elsewhere} \end{cases} \tag{15}$$

$$\begin{cases} \dfrac{\eta_{imm}^{[1][2]}(t_{ff}) - \eta_{imm}^{[1][2]}(t_{00})}{1 - \eta_{imm}^{[1][2]}(t_{00})} \dfrac{\Delta Q_i^{(j)[2]}}{t_{ff} - t_{00}} & \text{for} \quad t_{00} < t \le t_{ff} \\ 0 & \text{elsewhere} \end{cases} \tag{16}$$

$$\begin{cases} \dfrac{\eta_{imm}^{[1][2]}(t_{ff}) - \eta_{imm}^{[1][2]}(t_{00})}{1 - \eta_{imm}^{[1][2]}(t_{00})} \dfrac{\Delta n_r^{(j)[2]}\Big|_{pure}}{t_{ff} - t_{00}} & \text{for} \quad t_{00} < t \le t_{ff} \\ 0 & \text{elsewhere} \end{cases} \tag{17}$$

where $\eta_{imm}^{[1][2]}(t_{00}) = 0.6$ and $\eta_{imm}^{[1][2]}(t_{ff}) = 0.9$ by assumption. Applying the terms detailed in eqns (15) to (17) above results in the number of people in the recovered category for omicron being 92% of the number in the recovered category for delta on t_{ff} = 7 January 2022, which was judged to be sufficiently close to the aim of 90%.

6. Results

The model results are compared with data available on 5 January 2022. Fig. 2 shows the number of active infections projected by the model, compared against figures published by the Office of National Statistics. The match against all the available data is good, although this does not guarantee that the rest of the curve will be followed.

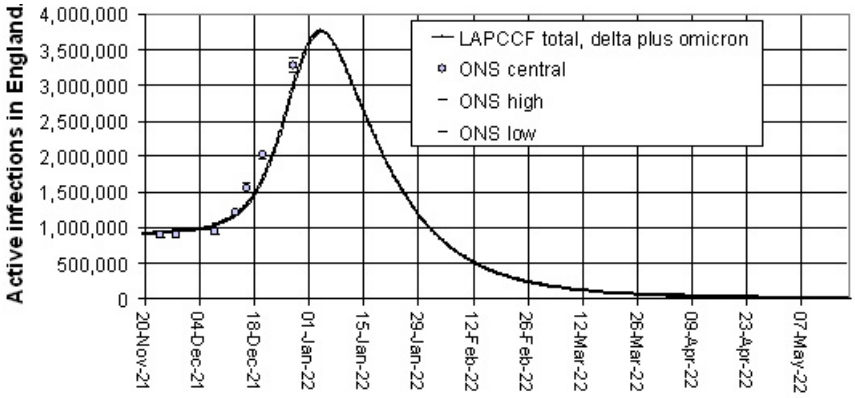

Figure 2. Active infections in England: model results compared against data from the Office of National Statistics Infection Survey (data available by 6 January 2022).

Fig. 3 shows the number of hospitalizations per day projected by the model for England, compared with official figures given by the Health Security Agency. Again the match is reasonable so far, although there is a lot of scatter in the real-world figures.

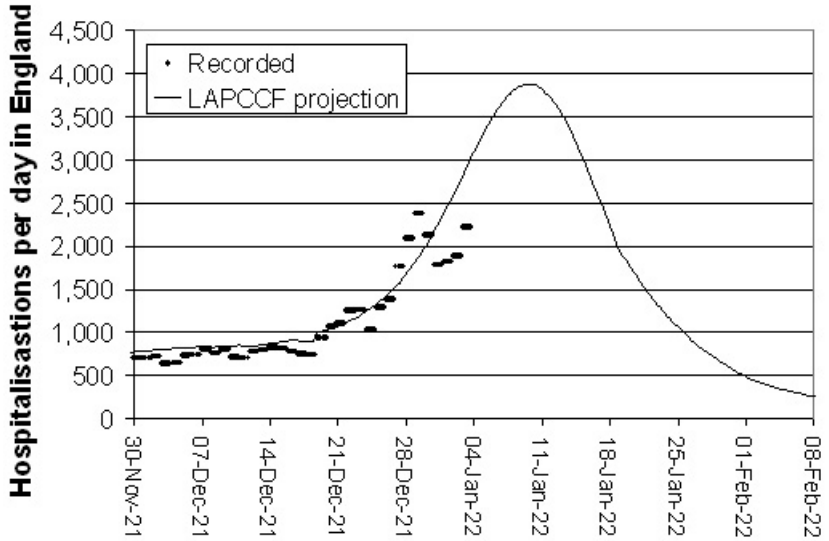

Figure 3. Hospitalizations per day in England: model predictions versus official figures from the Health Security Agency (data available by 6 January 2022).

Fig. 4 gives the projected deaths per day compared with figures from the Health Security Agency. So far the official data are not showing any significant upturn, although this might be a timing issue, with deaths being more delayed than allowed for.

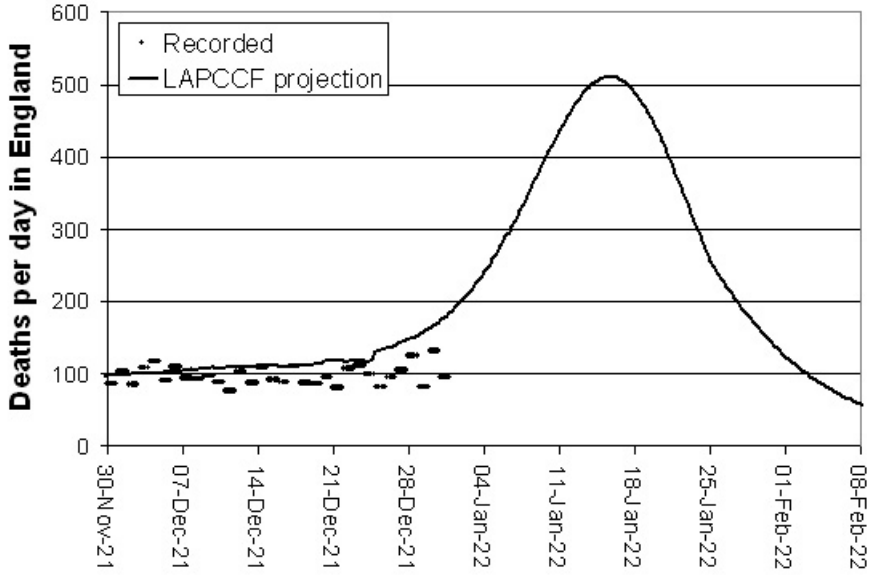

Figure 4. Deaths per day in England, model predictions versus official figures from the Health Security Agency (data available by 6 January 2022).

7. Sensitivity study

Putting the notional completion date t_{ff} for the booster campaign back 3 weeks to 28 January 2022 has the effect of increasing the peaks for the key parameters. The graphs are shown in Figs 5, 6 and 7. The peaks are about 60% higher. The match to the data to 5 January 2022 is generally not quite as good, but the mismatch does not appear to be sufficient to rule this scenario out of court.

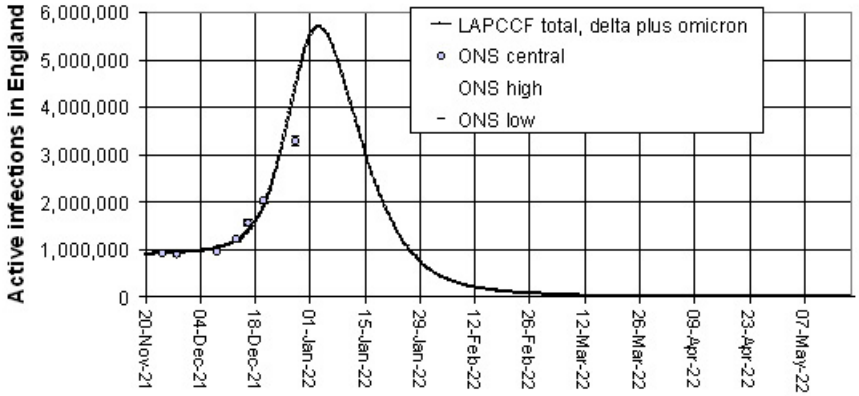

Figure 5. Active infections in England. Sensitivity study: putting t_{ff} back to 28 January 2022 (data available by 6 January 2022).

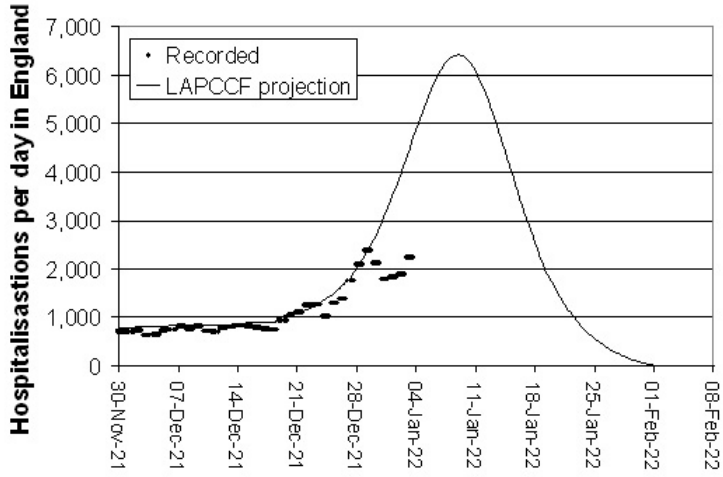

Figure 6. Hospitalizations per day in England. Sensitivity study: t_{ff} is put back to 28 January 2022 (data available by 6 January 2022).

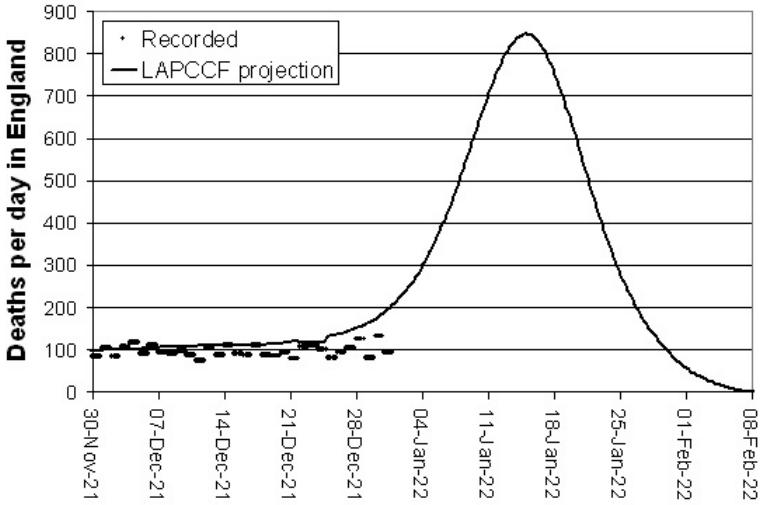

Figure 7. Deaths per day in England. Sensitivity study: t_{ff} is put back to 28 January 2022 (data available by 6 January 2022).

8. Discussion

The effect of the government's campaign of 3rd doses of vaccine has been modeled to highlight the likely main consequences of the spread of the omicron variant without restrictions beyond "Plan B" in England, under which people were told to continue working from home wherever they could, wear face coverings on public transport and in public places, and show their Covid status to get into nightclubs and large public events.

The model was kept relatively simple, in line with the aphorism attributed to the statistician George Box: "All models are wrong, but some are useful". The simple model is then compared with incoming data in order to test and validate it. The eminent philosopher Sir Karl Popper has told us that all theories and models retain their usefulness only for as long as they continue to be validated and not "falsified" by the data,[8] and the LAPCCF is no exception. At the time of writing, the incoming data have not falsified the model, certainly not as far as active infections are concerned (Fig. 2), nor, perhaps, as far as hospitalizations per day are concerned (Fig. 3).

The predicted peak for hospitalizations, just under 4000 per day, is slightly less than was seen in England in January 2021. Moreover, if the trend reported from South Africa is followed, hospital stays in England will be roughly halved. This has an important implication set down by the government's SAGE committee: "If the average length of stay with omicron were to be, for example, half as long as with delta on average, then for a given level of hospital admissions, occupancy would be approximately halved." High bed occupancy has the potential to cause problems for the National Health Service, but these should not occur if Covid bed occupancy peaks below half of last year's maximum.

The situation regarding deaths is less clear at the time of writing, since they stayed close to 100 per day throughout December 2021 (Fig. 4), and no clear upward trend has become apparent. It is possible that deaths will begin to rise in January 2022. On the other hand, it may also be that the omicron variant is milder than has been allowed for, which would keep deaths below the levels expected by the model.

The sensitivity study, whereby the full, beneficial effect of the booster campaign was delayed from early January to its end, shows active infections, hospitalizations per day and deaths per day all rising about 60% above their base case values (see Figs 5, 6 and 7). Hospitalizations peak at 6400 per day, but if hospital stays are halved this will be equivalent, in terms of bed occupancy, to only 3200 hospitalizations per day (in terms of last year's figures), which would not threaten the NHS.

9. Conclusions

Covid-19, in the many forms it has revealed itself to us, is inherently difficult to predict as the epidemic is naturally unstable, in the sense that, without active control, the numbers of infected people will either increase very rapidly (before saturating at a high number some way past the population immunity threshold) or else will collapse. However the simple scheme presented here has been able to explain some important features of the omicron epidemic in England, even while that epidemic is still progressing, as illustrated by Fig. 2 particularly. The peak of active infections is likely to occur early in January 2022. The base case model suggests that active infections will reach a maximum of roughly 3.5 million, but the 5.5 million figure suggested by the sensitivity study cannot be ruled out.

The model has proven to be a useful guide to the likely course of the epidemic. Moreover, the picture presented is generally hopeful for the future. There seems to be a good chance that England will emerge from this Covid wave without the NHS becoming overstretched.

---

[8] K.R. Popper, *The Logic of Scientific Discovery.* London: Hutchinson (1959; first published in Vienna in 1934 as *Logik der Forschung*).

Validation postscript: extra data available up to 4 March 2022

PS.1. Introduction

The model predictions detailed above, submitted to *Nanotechnology Perceptions* on 6 January 2022 using data available to 5 January 2022, were the basis for an article carried in the print edition of the weekly magazine *The Spectator*, dated 8 January 2022, with graphs (Figs 2, 3 and 4 of this paper) included in the online version.[9] An extra 2 months' of official data are now available and these are included in this Postscript.

No change has been made to the core of the model, which predicts active infections. Hence the extra data constitute a pure, unseen test of the validity of the LAPCCF's predictions for active infections, made on 6 January 2022.

However, later information from South Africa,[10] which was not used in the original model, suggested that hospitalizations per day for omicron would not be two thirds less than for delta, as reported in Abdullah et al. (2022)[1] and assumed in the model of 6 January 2022, but 80% down. Accounting for this newer information, the model was reprogrammed with the assumption that, while doctors began sending omicron-infected patients to hospital at a third of the rate they would have sent delta-infected people, this practice finished at the end of December 2021, from when on they hospitalized only a fifth, by comparison with delta. Their decisions on hospitalizations for patients with delta would, meanwhile, be taken in the same way as before.

Based on contemporaneous data on omicron and non-omicron Covid-19, Wolter et al. found that people, once hospitalized with omicron, given that their number was 80% down on the equivalent number who would have been sent to hospital with delta, now had the same chance of dying as before from that point on.[10] Applying this to England would suggest a death rate that would be $1/5 \div 1/3 = 3/5$ of the rate assumed in the model of 6 January 2022, where the omicron death rate was assumed to be a third of what it would have been with delta. This would have the effect of bringing the predicted peak death rate down from ~500 (Fig. 4) to ~300 per day. This change was also programmed into the updated model.

In addition, an effective R-rate across both variants, delta and omicron, was found using the method explained in Appendix PS.A, and compared against the R-rate derived at intervals by the React 1 study[11] and that derived on a weekly basis from the ONS Covid-19 infection survey.

These changes to the model were programmed on 8 January 2020.

PS.2. Comparison of the model against data available by 4 March 2022

PS.2.1. Comparison of model versus survey data for active infections in England

Fig. PS.1 shows the number of active infections against time in England as predicted by the LAPCCF and the survey measurements made by ONS and by React 1. It is clear that the LAPCCF predictions match the ONS data very well up to 12 January 2022, including capturing

[9] P. Thomas, Why the Omicron wave won't overwhelm the NHS. *The Spectator* (8 January2022).

[10] N. Wolter et al., Early assessment of the clinical severity of the SARS-CoV-2 Omicron variant in South Africa. medRxiv preprint (2021) doi: https://doi.org/10.1101/2021.12.21.21268116

[11] Real-time Assessment of Community Transmission (REACT) Study https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/real-time-assessment-of-community-transmission-findings/

both the high point and, importantly, the date when the numbers begin to decline. It should be emphasised that the vital, turnover point was not known at the time of the original submission of the paper to *Nanotechnology Perception*s (on 6 January 2022). The data from React 1, which reports less frequently than ONS and, arguably, has a slightly worse track record, are also matched reasonably well up to 12 January 2022. The fall in numbers measured by ONS between 12 and 18 January is seen to be somewhat less than predicted by the LAPCCF, and it is clear by 25 January that the ONS-measured active infections have leveled off. There is even an uptick in active infections in England over the week to 2 February. But a downward movement has set in by 9 February, and this falling trend continues over the next week to 16 February 2022.

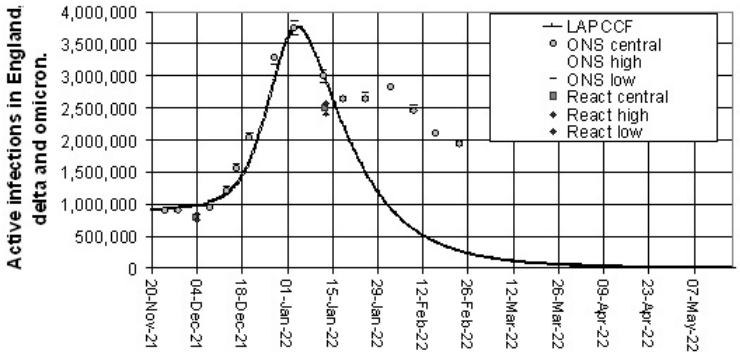

Figure PS.1 Active infections in England: model results compared against data from the Office of National Statistics infection survey and React 1 infection survey (data available by 4 March 2022).

An alternative perspective is provided by Fig. PS.2, which shows the LAPCCF R-rate compared against figures supplied by the React 1 survey and derived values based on the ONS surveys. The survey-measured R-rate, which was below 1.0 on 12 January, begins to rise thereafter, even exceeding 1.0 on 2 February before starting to fall again by 9 February 2022.

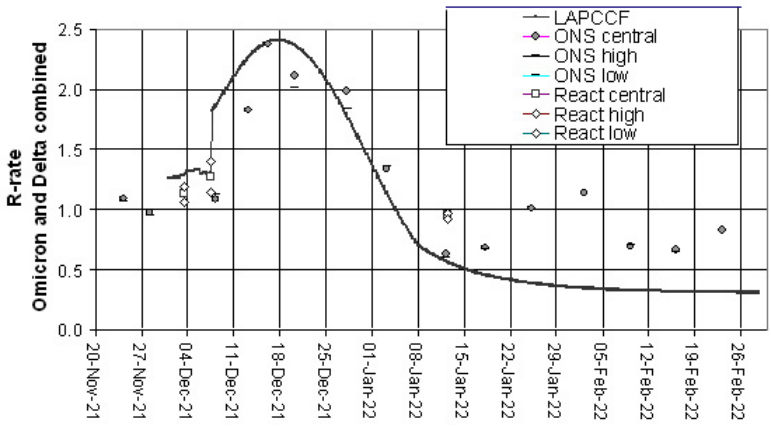

Figure PS.2 R-rate comparison between LAPCCF and React 1 survey figures and figures derived from ONS survey data (data available by 4 March 2022).

The temporary plateau in active infections, which was not predicted by the LAPCCF, is at least partly explicable in terms of the growth of a new omicron subvariant, BA.2. This unanticipated subvariant is clearly more infectious than the already extremely infectious original omicron subvariant, BA.1, because it is outcompeting BA.1, as shown by Fig. PS.3, which shows the increase of BA.2 relative to BA.1. (A Danish study, which analysed coronavirus infections in more than 8500 households between December 2021 and January 2022, found that people infected with the BA.2 subvariant were roughly 33% more likely to infect others, compared to those infected with BA.1.[12])

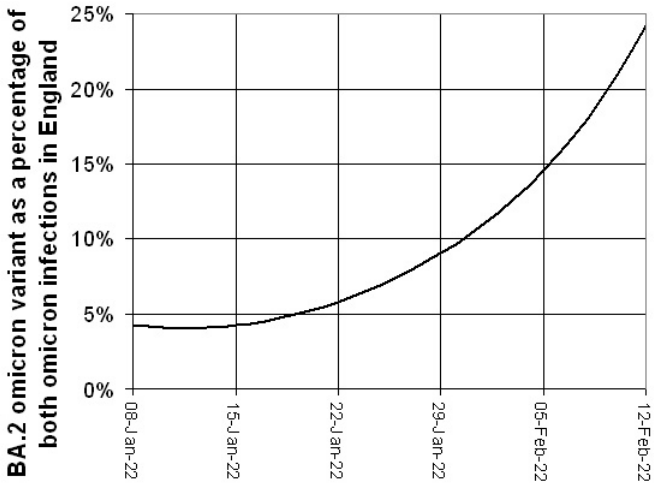

Figure PS.3 Prevalence of BA.2 omicron variant relative to original BA.1 variant (source: ONS Covid-19 infection survey of 16 February 2022).

The BA.2 variant began growing in prevalence at just the time when omicron's decline started to slow, around 15 January 2022. But while it was outcompeting BA.1 from that date, it would not have been able to stage a runaway expansion because of the very high level of immunity against all Covid-19 infections already present in the community. This immunity caused all active infections of Covid-19 to start falling again by 9 February 2022. It is likely that active infections will continue to fall at roughly the rates predicted by the LAPCCF, but between 2 and 3 weeks later.

PS.2.2. Comparison of LAPCCF model against UK Health Security Agency (HSA) data on hospitalizations per day

The LAPCCF, which now incorporates revised modeling of hospitalizations from 31 December 2021 onwards, shows reasonable agreement with the measurements until 22 January 2022 (once allowance is made for the pronounced weekly cycle characterizing recorded hospital

[12] Reuters, Omicron subvariant BA.2 more infectious than 'original', Danish study finds (31 January 2022) https://www.reuters.com/business/healthcare-pharmaceuticals/omicron-subvariant-ba2-more-infectious-than-original-danish-study-finds-2022-01-31/

admissions), see Fig. PS.4. The maximum recorded number of daily hospitalizations, 2370 on 29 December 2021, is close to the model projection of 2521, although the latter was predicted to occur 12 days later on 9 January.

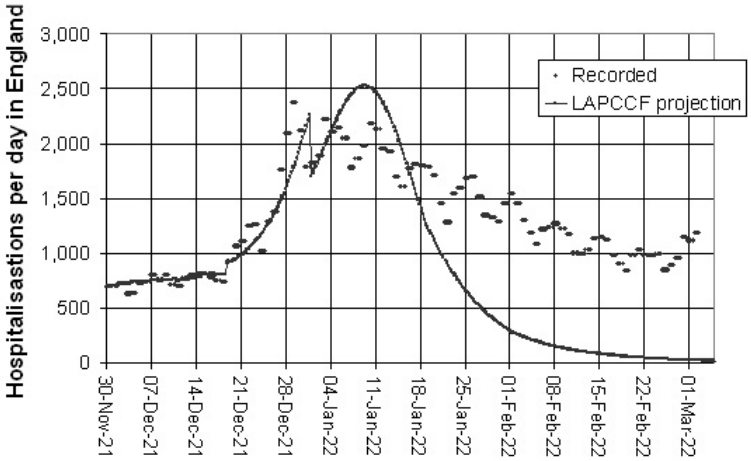

Figure PS.4 Hospitalizations per day: comparison of the LAPCCF model, with revised modelling from 31 December 2022, against UK Health Security Agency data published on the UK Government's Covid-19 Dashboard. (Data available by 4 March 2022).

The data after 22 January show a continuing overall fall but one that is slower than the LAPCCF model predicts. Daily hospitalizations were running around 1000 a day in mid-February, at a time when the model was predicting less than a tenth of that number. It is possible that doctors exercised greater caution in their decisions on whether or not to hospitalize a patient as the Covid-19 epidemic receded because they were aware that there would be no shortage of available hospital space.

PS.2.3. Comparison of LAPCCF model against recorded deaths per day

The possibility of subjective variation over time does not apply to recorded data on daily deaths (although it is accepted that the chosen statistic, deaths within 28 days of a positive test by date of death, still leaves room for argument on whether it represents deaths from Covid or with Covid).

As shown in Figure 4, which illustrates the data known at the time the paper was submitted on 6 January 2022, recorded deaths per day showed no material rise, by the end of December 2021, from the level they had been at for the previous month. However, the number of daily deaths started to rise immediately afterwards.

Figure PS.5 demonstrates that the LAPCCF model for daily deaths, incorporating the revisions discussed in §PS.1, gives very good agreement with the data up to 8 January, but the observed figures start to level off thereafter. The reduction in deaths may have been a beneficial consequence of the NHS introducing Pfizer's new oral Covid-19 antiviral drug, Paxlovid, which was licensed for use in the UK on 31 December 2021, and found, in trials with high-risk symptomatic adults, to reduce the risk of hospitalization and death within 28 days by 89%,

when the 5-day course of treatment was started within 3 days of symptom onset.[13] The first patients treated with Paxlovid would have started to complete their 5 day treatments towards the end of the first week in January, at very much the time when the difference between modeled and recorded death numbers began to become noticeable.

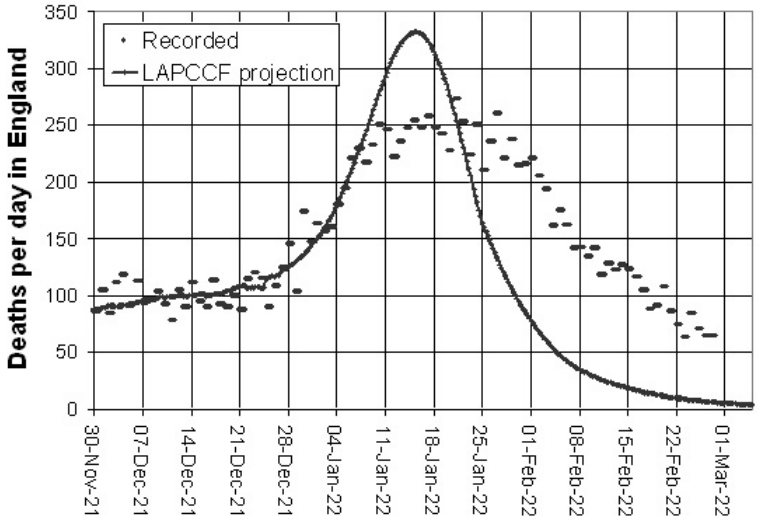

Figure PS.5 Deaths per day: comparison of the LAPCCF model, with revised modeling from 31 December 2022, against UK Health Security Agency data for deaths within 28 days of a positive test, published on the UK Government's Covid-19 Dashboard. (Data available by 4 March 2022).

Thus the LAPCCF, even with the revisions incorporated, overestimates the peak in deaths by about 20% (331 on 15 January predicted, 273 on 21 January registered).

Recorded deaths per day are clearly falling from about 28 January, roughly 12 days after the model predicts, and the downward trend continues from then on, dropping to about 100 per day by 17 February 2022. This delay between modeled and recorded deaths per day falling is similar to what was seen with modeled and survey-measured active infections, which suggests that the prolonged leveling off of deaths observed is likely, once again, to have been at least partly due to the additive effect of the more contagious BA.2 omicron subvariant.

Deaths per day seem now to be following the trend suggested by the LAPCCF, but, again, about two weeks later. They had already fallen to 75 per cent below their January peak by the end of February 2022.

PS.3 Discussion

It is clear that the original LAPCCF model, as described in the paper submitted to *Nanotechnology Perceptions* on 6 January 2022, gave accurate predictions of both the peak

[13] Medicines and Heathcare products Regulatory Agency (MHRA), Oral COVID-19 antiviral, Paxlovid, approved by UK regulator (press release 31 December 2021) https://www.gov.uk/government/news/oral-covid-19-antiviral-paxlovid-approved-by-uk-regulator

number of active infections in England and the date on which the peak would occur. No alterations were programmed into this part of the model, and so this good result is preserved in the updated model of 8 January 2022.

The original model's subsidiary predictions, of hospitalisation per day and deaths per day in England, gave peaks that were too high: 3879 (cf. 2370 recorded) and 511 (cf. 273 recorded) respectively. Therefore revised and improved calculations for these variables were programmed into the LAPCCF model on 8 January 2022. Peak predicted hospitalizations per day fell to 2521 as a result, reducing the error between predicted and actual to 6%. The recorded maximum number of deaths per day, 273, was 18% less than the 332 modeled peak, but this welcome reduction may have been brought about, at least in part, by the NHS starting to use Pfizer's powerful new Paxlovid antiviral drug after it was licensed by the MHRA on 31 December 2021.

The later data available to 4 March 2022 showed falls in active infections, hospitalizations and deaths delayed by two or so weeks compared with the later LAPCCF predictions. The overall picture of peaks in early to mid January followed by prolonged falls was retained. Moreover, it looks likely that the two-week delay observed in the actual data was caused by the growth of the new more infectious BA.2 variant, which started to take off in the middle of January 2022.

The validation against unseen, new data confirms that the LAPCCF model, either in its original or improved form, gave predictions, both for the key variable of active infections and for the subsidiary variables of hospitalisations and deaths per day, that were sufficiently accurate to be useful as a guide to policy.

The success of the LAPCCF contrasts with the contemporaneous but poorer performance of the models programmed by the scientists reporting to the government's Scientific Advisory Group on Emergencies (SAGE), as analysed in detail by Michael Simmons.[14,15]

Graham Medley, Chairman of SAGE's subcommittee on modeling, SPI-M, attempted to defend the performance of the modeling groups reporting their findings to government.[16] Prof. Medley said that they were encouraged to model a range of scenarios, but admitted that "in the SPI-M scenarios for the summer of 2021 after the lifting of restrictions ... real Covid hospitalizations and hospital occupancy are, thankfully, currently at the bottom end of the range of scenarios we put forward for what might happen". He added, moreover, that "SPI-M produced consensus scenarios at the start of September looking at the potential impact of transmission in schools. Again, numbers are below the lower scenario".

In fact, actual numbers of cases, hospitalizations and deaths have shown a consistent tendency to come out below the lowest scenario put forward by SPI-M. Overestimates by SAGE lie at the heart of the SPI-M-O paper, "Long term winter scenarios preparatory working analysis" of 31 October 2020, for example. The best estimates of deaths per day without a lockdown for 4 November 2020, the day a new lockdown was introduced as a result of SAGE's advice, were 310 (Warwick), 380 (London School of Hygiene and Tropical Medicine (LSHTM), 570 (Imperial College) and 1220 (Public Health England/Cambridge), when the actual figure was 278.

Moreover, in a revealing Twitter exchange with Fraser Nelson, on 18 December 2021, Prof. Medley appeared either unable to grasp, or at least to question, the point of modeling

[14] M. Simmons, SAGE scenarios vs actual: an update. *The Spectator* (16 January 2022).
[15] M. Simmons, SAGE scenarios vs actual: an update. *The Spectator* (23 January 2022).
[16] G. Medley, In defence of SAGE's models. *The Spectator* (22 October 2021).

scenarios that did not show outcomes bad enough to oblige politicians to impose extra restrictions.[17] Answering Nelson's comment: "I guess the question is why LSHTM[18] did not (like JP Morgan) include a scenario of lower virulence—given that this is a very-plausible option that changes outlook massively", Medley said, "What would be the point of that? Not a snarky question—genuine to know what you think decision-makers would learn from that scenario." Nelson: "In the low-virulence scenario modeled by JP Morgan, no further restrictions would be needed so harm to economy and society might be averted. Can I ask why you didn't think this less alarming (and quite plausible) scenario was worth including? Like yours, a genuine question ..." Medley: "You know the answer. That's what the paper says. If somebody draws a line on a graph it doesn't add any further information.[19] Decision-makers are generally only interested in situations where decisions have to be made". Nelson: "I may be being thick but I'm afraid I don't know the answer! Why would you not—for completeness—add the scenario where Omicron is less virulent and more restrictions are not needed?" Medley: "I meant you know what happens. The scenario does not inform anything. Decision-makers don't have to decide if nothing happens". Nelson: "Thanks, this helps me understand. So you exclusively model bad outcomes that require restrictions and omit just-as-likely outcomes that would not require restrictions?" Medley: "We generally model what we are asked to model. There is a dialogue in which policy teams discuss with the modelers what they need to inform their policy". Nelson: "Okay, so you were asked to model bad Omicron outcomes and make no comment as to the probability?" Medley: "We model the scenarios that are useful to decisions".

Concentrating almost exclusively on worst case or even "reasonable worst case" scenarios, if this is what the epidemic modelers decided to do or were expected to do, would represent a break from normal modeling practice as applied to safety analysis of major hazards in other sectors. In the examination of nuclear reactor transients, for example, it is normal to use best-estimate parameter values to make a central prediction of the dynamic behaviour of key variables. The results will constitute the modeler's best prediction of what is most likely to happen using the data in which people can have most confidence. Sensitivity studies are then used to examine less likely scenarios in which parameters may deviate from their expected values. Ideally, probabilities will be assigned to the resultant, more unlikely, outcomes and the decision-maker will be presented with a central, most likely, outcome together with a less probable pessimistic outcome and, perhaps, a less likely optimistic outcome to give balance.

This fits in with the normal way of presenting measurements that are subject to random variations, where a most likely best estimate is given, together with a 95% confidence interval. Such an approach is adopted by the ONS, for example, when it provides its weekly estimates of active infections in England, based on surveys of 150 000 people, which are regarded as the gold standard.

---

[17] F. Nelson, My Twitter conversation with the chairman of the SAGE Covid modelling committee. *The Spectator* (18 December 2021).

[18] Prof. Medley is affiliated with the London School of Hygiene and Tropical Medicine.

[19] This assertion seems strange, given the centrality of "lines on graphs" to the entirety of science and mathematics. Prof. Medley may have been thinking of a horizontal straight line on a graph of a variable against time, which would mean that the variable was constant over the interval. But the knowledge that something which might vary did not, in fact, change over an interval carries significant meaning and may, indeed, be rather important.

If we imagine that the worst-case modelling scenarios provided by SAGE reside at one extreme of a 95% confidence interval, then a decision to use these as the basis for setting policy implies that the policy-maker will have spent his or her time devising countermeasures appropriate to outcomes that, we can assert, with 97.5% confidence, will be worse than whatever was destined actually to happen without intervention.

Such a strategy can only work if the precautions are relatively painless. But this has never been the case with Covid restrictions, which have closed down society and the economy, with delayed, knock-on effects on health that were predicted scientifically[20,21] in March 2020 and which were starting to become more widely realized and acknowledged[22] by late 2021.

The probability of the situation playing out without intervention as badly or worse than a worst-case scenario may well be much lower than 2.5% if the worst case scenario is constructed from the concatenation of a series of pessimistic estimates of the values of input parameters. The resulting lack of realism has been explained clearly by Fremlin who, in his book on the risks associated with power production,[23] states: "I will begin by saying that I shall not be concerned with the maximum credible accident. This depends on the imagination of the inventor and the credulity of the receiver—neither of which has any definable limit". He cites, as an example, an air crash above London:

> It is perfectly *possible* for a couple of off-course, fully-fueled Jumbo jets to collide, one of them crashing on Wembley stadium during a cup final, bits of it blocking the exits, and a hundred tons or so of kerosene—a litre each for 100,000 people—being scattered over the crowd and ignited; while the second Jumbo crashes on Canvey Island destroying the LNG storage tanks and a refrigerated supply vessel on the Thames just coming in with fresh supplies, all at a time when there is a suitable breeze to blow the cold gas down into Canvey Town, with its 30,000 inhabitants, before it explodes. There is more than enough LNG stored there to kill the whole population of Canvey and burn the entire town to the ground. Then a fire engine dashing to the spot collides with a chlorine tanker ... and so on and so on.
>
> ... this is actually possible, merely unlikely.
>
> The fact is that no decisions can be made on the basis of the worst possible case, and in our own lives we never do allow such cases to affect our plans.

It might be argued that the "reasonable worst-case scenario" and the worst possible case are different things, but in the absence of explicit probability statements what methodological rule can be applied to ensure that the former does not, in practice, degenerate into the latter? An important practical point, especially in new situations (such as an epidemic of a new disease like Covid-19) where parameters are difficult to estimate, is that it is much easier to identify a good central estimate of a parameter's value than to define its probability distribution. The best

[20] P. Thomas, J-value assessment of how best to combat COVID-19. *Nanotechnol. Perceptions* **16** (2020) 16–40.
[21] T. Whipple, Economic crash could cost more life than coronavirus, says expert. *The Times* (24 March 2020).
[22] S. Knapton, Thousands more people than usual are dying ... but it's not from Covid. *Daily Telegraph* (24 September 2021).
[23] J.H. Fremlin, *Power Production: What are the risks?*, pp. 123–124. Oxford: University Press (1987).

estimate prediction derived from central parameter estimates is likely, as a result, to be a more dependable statistic than the confidence interval. While the formulae needed for calculating the confidence interval may be available, the probability distributions on which the mathematics has to operate are likely to be rather poorly characterized.

Providing the decision-maker with a small number possible outcomes, as is the case with the ONS measurements of active infections, gives him or her a reasonable chance of taking a well-informed decision, but the necessary clarity is likely to break down when the modeler passes on a very large number of possible outcomes. The presentation of multiple scenarios, especially without a comment on their likelihood, may, of course, provide the modelers with the defence (against tests of their predictions—or "projections"—against real data) that their lack of knowledge has been declared at the outset. But it is difficult, for example, to see how a policy-maker's judgment will be improved when presented with twenty graphs, as was the case for hospital admissions contained in Annex 1 of the "SPI-M-O Consensus statement on COVID-19" of 6 January 2022, where each graph was itself made up of multiple curves that attempted to illustrate the effect of different policy options.

The government's Chief Scientific Adviser has argued that "It's not true COVID-19 modellers look only at worst outcomes",[24] but Sir Patrick Vallance's assurances seem to sit uneasily with the comments made only the week before by the chairman of SAGE's modeling subcommittee. Indeed, the historic record of the projections made by the modelers advising the government looks rather worse if they were intended as best-estimate predictions of what was most likely to happen. SAGE modelling committee chairman, Prof. Medley, later stated to a House of Commons select committee his belief that Covid modeling could not accurately predict numbers. Speaking on 2 March 2022 at the Science and Health Select Committee, which was scrutinizing the pandemic response, Prof Medley said: "The modelling is there to understand the process and what's going on. We know we can't accurately predict the numbers but we can give insight into the processes that determine the outcomes".[25]

In fact, the modelers reporting to SAGE in mid-December 2021 regarded 3000 hospitalizations and 600 deaths per day as the minimum that would be incurred without restrictions beyond Plan B. Furthermore, the SPI-M modeling subcommittee suggested that the hospitalizations could rise to 10 000 per day, roughly two and a half times England's previous peak of 4,134 recorded on 12 January 2021, while deaths could rise to 6000 per day respectively, nearly five times the highest level, 1249, seen in England on 19 January 2021.[26] SAGE modellers said that the only way to ensure hospitalizations per day were brought below 5000 per day and deaths below 2000 a day (both of which figures would still constitute the highest ever seen, by some margin) would be to reimpose restrictions for three months from January to March 2022 that would be equivalent to England's Step 1 out of lockdown the year before.

[24] P. Vallance, It's not true COVID-19 modellers look only at worst outcomes, published in *The Times* on 24 December 2021 and on the government's website on the same day, https://www.gov.uk/government/speeches/its-not-true-covid-19-modellers-look-only-at-worst-outcomes

[25] S. Knapton, Covid modelling cannot accurately predict numbers, admits government expert, *Daily Telegraph* (2 March 2022).

[26] SPI-M-O: Consensus Statement on COVID-19, 15 December 2021, Table 1, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1042204/S1439_SPI-M-O_Consensus_Statement.pdf

Step 1, which was introduced on 8 March 2021, it will be remembered, retained the strict, two-month lockdown imposed on 5 January 2021, where people were not allowed by law to leave or be outside of their homes without a "reasonable excuse",[27] but with the relaxations (i) that people would be allowed to leave home for recreation and exercise outdoors with their household or support bubble, if they were eligible for one, or with one person from outside their household, (ii) that care home residents would be allowed one regular visitor[28] and (iii) that schools would be allowed to open.

In the event, no extra restrictions were imposed beyond Plan B, yet the actual peak number of hospitalisations, 2521 per day, was only a quarter of SAGE's top figure, while the maximum number of deaths in England, 273 deaths per day, was a staggering 95% less than the highest figure that SAGE regarded as plausible.

It is, perhaps, hardly surprising, given its history of inaccuracy, that SAGE was stood down on 4 March 2022, with the *Daily Telegraph* reporting that the "Government's scientific advisory group has faced criticism for modelling, which has been repeatedly shown to be wrong", and that "Behind the scenes, the *Daily Telegraph* understands ministers have become increasingly frustrated with pessimistic predictions from SAGE and are concerned the group does not fully take into account the damage to the economy, mental health and education from lockdown".[29]

SAGE's record on modeling has been so poor that some have wondered whether models have much use at all, with Prof. Carl Henegan, who is Professor of Evidence-Based Medicine and Director of the Centre for Evidence-Based Medicine at the University of Oxford, quoted in the *Daily Telegraph* article just cited as saying that SAGE was overreliant "on modelling—which we now know is no more than 'guesswork' ", while his colleague Dr Tom Jefferson reinforced the point by commenting: "The work of this group was all about guesswork, and if you're going to be ruled by guesswork then we're all doomed. The modellers have a track record of getting things wrong".

Yet it is clear from the validation, presented in this Postscript of the LAPCCF, that a model that was mathematically equivalent to Kermack and McKendrick's 1927 dynamic equations of epidemic behaviour[30] (the SIR model) was indeed able to give a good best-estimate prediction of the course of the Covid-19 epidemic as the omicron variant started to dominate. It has been demonstrated how modeling can, indeed, be useful, and it makes sense to explore how the LAPCCF achieved its good results.

The first thing to observe is that the LAPCCF is a best-estimate model. This simpler approach appears to have been justified by the accuracy of the predictions. Then it should be noted that the LAPCCF applied its equations to England as a whole, without resort to regional

---

[27] South Somerset District Council, COVID-19: Government announces new period of national lockdown from Tuesday 5 January (4 January 2021) https://www.southsomerset.gov.uk/news/2021/1/covid-19-government-announces-new-period-of-national-lockdown-from-tuesday-5-january/

[28] Cabinet Office, COVID-19 Response - Spring 2021 (Summary), Roadmap out of lockdown (22 February 2021) https://www.gov.uk/government/publications/covid-19-response-spring-2021/covid-19-response-spring-2021-summary

[29] S. Knapton, Sage stands down, signifying 'end of Covid pandemic' in the UK. *Daily Telegraph* (4 March 2022).

[30] W.O. Kermack and A.G. McKendrick, A contribution to the mathematical theory of epidemics. *Proc. R. Soc.* **A 115** (1927) 700–721.

disaggregation. It seems, therefore, that the extra complication added by splitting the country up into regions, applying the equations regionally and then recombining the results to give the national picture (something that at least some SAGE models did) was unnecessary.

However, the LAPCCF did divide England's population of 56.5 million people into 8 age groups: 0–24, 25–49, 50–54, 55–59, 60–64, 65–69, 70–79 and over 80, which allowed for different levels of vaccination and recovery in each age interval. Moreover, the people in each age group were further divided into two symptom cohorts: those who experienced symptoms after contracting Covid-19 and those would be either asymptomatic or show only minor symptoms.

While allowance needs to be retained for Covid's unusual characteristic that a significant proportion of the people it infects never experience symptoms, it may be that the LAPCCF's disaggregation of the population into 8 age groups is not necessary. Sufficient accuracy might be achieved with a smaller number of age groups, possibly even modeling the population as a single group, without regard to age, which was the approach used with the original PCCF. Simpler models are less prone to error and give greater flexibility for "optioneering" or the assessment of different policy options, and tend to give results that are easier to explain. They also offer a valuable, diverse check on the results that might come from more complex models. (Engineers needing to assure their designs often carry out such diverse checks using simplified analytic models,[31] but the Kermack and McKendrick equations, while simple in form,[32] do not admit of an analytic solution.) There is a good case to be made for policy makers always to ask for a check of the results from a complex model against those produced by the simplest possible model, with an explanation sought of why any significant differences arise.

As explained in detail in §4, the LAPCCF was matched to UK HSA and ONS data on Covid-19, up to 7 December 2021, by adjusting key parameters. Benefit was then taken of the fact that, to a good approximation, all the infections were delta until this point. The LAPCCF was able to extrapolate the number of new delta infections there would be each day, up to about 10 days into the future. The difference between these extrapolated case numbers for delta and the official figures for all new cases each day would then constitute a good estimate of the new daily omicron cases.

The method was made possible by the experience that had been built up of matching the LAPCCF to the epidemic on a daily basis for the previous 6 months and the PCCF for over a year. This established the practicality of gathering near real time information on the key R-rate variable (which was then published daily, 4 days in arrears, on *The Spectator*'s data hub for over 12 months, for information and as a potential guide for controlling the Covid-19 epidemic). This rapidity contrasts with the delayed availability of SAGE's estimate of the R-rate, which was based on a consensus amongst modeling groups, and was published only once a week, by which time it was admitted already to be "2 to 3 weeks" in arrears,[33] making SAGE's R-rate figure of historic interest at best. Effective control of a fast moving epidemic like Covid relies on a rapid measurement of the R-rate, and the fact that the government's

[31] A detailed example is given in P.J. Thomas, T.A. Harrison and P.D. Hollywell, Analysis of limit cycling on a boiler feedwater control system. *Proc. 3rd International BNES Conference on Boiler Dynamics and Control in Nuclear Power Stations*, Harrogate, 21–25 October 1985.

[32] A derivation of equations equivalent to the Kermack and McKendrick model[30] is given in Appendix A. Model of the Covid-19 epidemic in the UK, of Thomas' 2020 paper.[20]

[33] UK Health Security Agency, Section "Time delay of the estimates" in *Guidance: The R value and growth rate*, https://www.gov.uk/guidance/the-r-value-and-growth-rate, update of 25 February 2022.

advisers were content for this key parameter to become known only after a long time delay reveals an astonishing lack of ambition.

While modeling can and should produce information of scientific interest, policy makers need to be concerned not so much with "following the science" as "engineering the solutions". Take vaccination, for example. While the science of genetics provided us with the makeup of the SARS-Cov-2 virus early on, this was only the first step on the road to producing an effective vaccine, where the vaccine developers and manufacturers needed to build on the science to engineer a solution by designing, producing and distributing an effective vaccine. Epidemic models, if they are to be useful in controlling an epidemic like Covid-19, need to be updated daily to allow decision-makers to have the earliest possible notification of the state of the epidemic that they are seeking to control.

The rapid estimation of the dynamic variables characterizing the Covid-19 epidemic depended on reconciling the cases by date reported, which were supplied daily by UK Health and Security Agency (HSA) up to the end of January 2022 (and earlier by its predecessor body, Public Health England, PHE) with the survey figures for active infections in England, which the ONS published weekly. The process relied on normalizing each day's positive cases to a standard number of tests so as to achieve comparability from day to day, even when the number of tests taken rose by an order of magnitude over the first year. The tests administered daily were originally of the PCR type and were given to people with symptoms that they felt might be Covid-19. It was possible, in these circumstances, to devise a reasonably accurate procedure for normalizing the results to a standard number of daily tests, in spite of the non-random, self-selecting nature of the reporting.

However the government decided later to introduce compulsory lateral flow tests for all school children, whether or not they had experienced any symptoms. The selection of whole age groups for test in this way was very different from the original process for choosing who would be tested, but the authorities nevertheless bundled all positive results together into the same category of cases by date reported, which was the basic statistic that needed to be used if the estimation delay was to be kept to no more than 4 days. A way of coping with this was again found, but it was necessarily more approximate.

A further complication of the PHE/ UK HSA reporting process was that, although new daily cases of Covid-19 were reported each day, the testing figures, which were essential to make sense of the numbers of cases, were not published over weekends nor on public holidays: every weekend brought a data desert, with the numbers of tests taken on Friday, Saturday and Sunday not arriving until the following Monday. Again, ways around this were found, but the accuracy of the readings was further compromised.

It is not clear why, when new cases were reported every day, it should not have been possible to publish the number of tests, at the same time. Clearly this ought to happen if and when England is hit by a similar epidemic in the future. Attention should also be paid to categorizing the data on new cases in such a way as to facilitate rapid and accurate comparisons between the numbers of new cases on two different dates and so allow the most rapid characterization of the state of the epidemic each day. Control engineers know that rapid measurement is a *sine qua non* for good control of any process, and that long measurement delays spell doom for any control system.

PS.4. Conclusions

The extra data that has come in since the initial submission of the manuscript on 6 January and by the time of writing on 4 March 2022 have validated the LAPCCF and its assumptions as applied to the dynamic behaviour of active infections in England. The advent of the BA.2 subvariant of omicron may have been a major factor in causing the drop in active infections to stall during the second half of January 2022 before they resumed their long fall, as predicted by the LAPCCF.

The revisions to the model in the area of hospitalizations per day and deaths per day have also allowed good predictions to be made for these output variables. The fact that the peak in recorded deaths per day is 18 per cent lower than the LAPCCF prediction may be explained, at least in part, the introduction of the highly effective Paxlovid antiviral drug for high-risk adults after it received its UK licence on 31 December 2021. The subsequent, roughly two-week delay in the fall in recorded compared with predicted deaths was occasioned, at least in part, by the growth of the BA.2 omicron subvariant.

The predicted peak in hospitalizations per day gets to within 6% of the actual number. Unlike the cases of active infections and deaths per day in England, the fall in actual hospitalizations per day shows no particular plateau in late January, but the rate of decline is generally less steep than predicted. It is possible that doctors are seeing the opportunity provided by continuing low levels of Covid-19 bed occupancy to be more cautious and thus to send higher numbers of less-sick patients to hospital than otherwise.

The LAPCCF has been validated as far as its predictions for the key variables are concerned: active infections in England, hospitalizations per day and deaths per day, in the sense that its results have been close enough to what actually happened to constitute a useful guide to policy.

The Discussion of §PS.3 explores how the LAPCCF was able to give good predictions of the course of the Covid-19 epidemic in England after the omicron variant arrived. Suggested lessons for the future have been sketched out. In particular, it is suggested that

i. it was possible to model the future course of the Covid-19 epidemic accurately, and there is no reason why an epidemic of a new disease should not be modeled accurately in the future;

ii. epidemic models, if they are to be useful in controlling an epidemic like Covid-19, need to be updated daily to allow decision makers to have the earliest possible notification of the state of the epidemic that they are seeking to control;

iii. most emphasis should be placed on best-estimate epidemic modeling as a guide to government policy;

iv. division of the population into regions for modeling purposes is unlikely to be necessary for accurate prediction;

v. if more complex models are retained, their results should always be checked for accuracy against a simple model, which would not split England into regions and might not need to divide the population into different age groups; discrepancies between the results from the complex models and from the simple model should be explained;

vi. it has been demonstrated that it was possible, by incorporating the modeling equations into a measurement filter, to use the data provided daily by the UK HSA and weekly

by ONS to measure key dynamic variables with a delay of no more than 4 days—roughly the time between infection and first onset of symptoms;

vii. these rapid measurements could have been used to assist with near-real-time control of the Covid-19 epidemic in England;

viii. the rapid estimation of key epidemic variables relied heavily on the daily record of "cases by date reported" and the reported numbers of tests administered daily, as published by UK HSA (and earlier by PHE). "Daily" should, however, mean just that, and the relevant data need to be supplied every day, including weekends and bank holidays;

ix. the weekly information provided by ONS infection survey proved very useful, but means of reducing the delay between measurement and publication should be explored; moreover, weekly needs to mean weekly, with no time-outs over the key Christmas period, for example;

x. heuristic algorithms were required to establish a level playing field as the number of tests taken per day increased rapidly because of the growing availability of tests, and the difficulty of normalizing the cases to a standard testing régime was compounded when lateral flow tests were made compulsory for school children, and when results from very different testing régimes were bundled together. While a reasonable degree of accuracy was retained, as exemplified by the validation of the modeling, more attention should be paid in future to designing the way in which daily data are captured and presented to allow more near-real time modeling.

Appendix PS.A. Averaging the R-rate over omicron and delta variants

The R-rate is the average number of people someone infected with Covid-19 will infect. It influences the dynamics of the spread of the virus through equation (PS.A.1), which allows the number of daily new infections, $\mathrm{d}n_x^{[k]}/\mathrm{d}t$, to be calculated, where $k = 1$ signifies the delta variant while $k = 2$ denotes the omicron strain:

$$\frac{\mathrm{d}n_x^{[k]}}{\mathrm{d}t} = \frac{R^{[k]}}{\tau_{\mathrm{inf}}^{[k]}} n^{[k]} \qquad k = 1, 2. \tag{PS.A.1}$$

Here $R^{[k]}$ is the R-rate for variant k, $n^{[k]}$ is the number of people infected (and infectious) with variant k, while $\tau_{\mathrm{inf}}^{[k]}$ is the average time between infections.

The number of people $\Delta n_x^{[k]}$ who are infected with variant k between epochs t and $t + \Delta t$, is given by

$$\Delta n_x^{[k]} \approx \frac{\mathrm{d}n_x^{[k]}}{\mathrm{d}t} \Delta t \,. \tag{PS.A.2}$$

The probability $p^{[k]}$ that someone who is infected with Covid between epochs t and $t + \Delta t$ has strain k is thus given by

$$p^{[k]} = \frac{\Delta n_x^{[k]} \Delta t}{\sum_{k=1}^{2} \Delta n_x^{[k]} \Delta t} = \frac{\Delta n_x^{[k]}}{\sum_{k=1}^{2} \Delta n_x^{[k]}} \approx \frac{\mathrm{d}n_x^{[k]}/\mathrm{d}t}{\sum_{k=1}^{2} \mathrm{d}n_x^{[k]}/\mathrm{d}t}. \tag{PS.A.3}$$

Thus the average or expected value, $E(R)$, of the R-rate is given by:

$$E(R)=\sum_{k=1}^{2} p^{[k]}R^{[k]} \approx \frac{\sum_{k=1}^{2} R^{[k]}\mathrm{d}n_x^{[k]}/\mathrm{d}t}{\sum_{k=1}^{2} \mathrm{d}n_x^{[k]}/\mathrm{d}t}. \tag{PS.A.4}$$

Eqn (PS.A.4) gives the degenerate solutions:

$$E(R)=\begin{cases} R^{[1]} & \text{before omicron appeared and only delta was around} \\ R^{[2]} & \text{once omicron becomes completely dominant} \end{cases} \tag{PS.A.5}$$

which matches what one would expect.