*Journal of the Royal Statistical Society Series A: Statistics in Society*, 2026, **189**, 1–31
https://doi.org/10.1093/jrsssa/qnaf092
Advance access publication 1 July 2025
**Discussion Paper**

ROYAL STATISTICAL SOCIETY DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
**Series A**
Statistics in Society

# Some statistical aspects of the Covid-19 response

**Simon N. Wood[1], Ernst C. Wit[2], Paul M. McKeigue[3], Danshu Hu[1], Beth Flood[1], Lauren Corcoran[1] and Thea Abou Jawad[1]**

[1]School of Mathematics, University of Edinburgh, Edinburgh, UK
[2]Institute of Computing, Univerità della Svizzera italiana, Lugano, Switzerland
[3]College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK

*Address for correspondence*: Simon N. Wood, School of Mathematics, University of Edinburgh, Edinburgh EH9 3FD, UK. Email: simon.wood@ed.ac.uk

Read before The Royal Statistical Society at the Discussion Meeting on 'Some statistical aspects of the Covid-19 response' held at Hallam Conference Centre in London on Thursday, 10 April 2025, the President, Professor Sir John Aston, in the Chair.

## Abstract

This paper discusses some statistical aspects of the UK Covid-19 pandemic response, focussing particularly on cases where we believe that a statistically questionable approach or presentation has had a substantial impact on public perception, or government policy, or both. We discuss the presentation of statistics relating to Covid risk, and the risk of the response measures, arguing that biases tended to operate in opposite directions, overplaying Covid risk and underplaying the response risks. We also discuss some issues around the presentation of life loss data, excess deaths, and the use of case data. The consequences of neglect of most individual variability from epidemic models, alongside the consequences of some other statistically important omissions, are also covered. Finally, the evidence that full stay-at-home lockdowns were necessary to reverse waves of infection is examined, with new analyses provided for several European countries.

**Keywords:** Covid models, excess deaths, lockdown, risk management, risk presentation, transmission heterogeneity

## 1 Introduction

Covid-19 caused immense strain on health systems and societies worldwide with the WHO official death toll to date[1] —generally considered a lower bound—corresponding to almost 0.1% of the world population or a life loss of around 3 days per capita. Although governments around the world undertook strenuous measures to mitigate the threat, these did not come without side effects. The UK Covid response caused substantial collateral damage. Directly in terms of blocked or delayed access to healthcare (e.g. Riera et al., 2021), exacerbation of mental health problems (e.g. O'Connor et al., 2021), lost schooling and normal social development for children (e.g. Major et al., 2021), and all the other human cost of cutting the great majority of the population off from normal social contact for months, under regulations that at times confined people to their own homes for 23 hr per day and prohibited outside contact other than online. Indirectly, and perhaps more substantially, through the human effects of the economic disruption, which the bank of England estimated was the largest for some 300 years, for the UK, and was unprecedented in that it involved deliberate halting of much economic activity, with money creation being employed to attempt to mitigate the consequent immediate problems. Creating money while reducing real

[1] Time of writing May 2023.

**Received:** June 7, 2023. **Revised:** December 15, 2024. **Accepted:** December 16, 2024

economic activity is obviously inflationary (e.g. Hall, 2009; Wolf, 2023, Ch. 1) even without the supply chain problems (e.g. Hutton & Powell, 2021) that followed on the resumption of paused activity.[2] The subsequent sharp increase in inflation[3] is one path by which the disruption has contributed to increased economic deprivation (e.g. Richardson et al., 2023; Shine, 2022) of the sort clearly linked to substantially reduced life expectancy and quality of life (e.g. Marmot et al., 2020).

Some proportion of these effects would have happened under any response to Covid, and they obviously have to be considered against the reduction in human suffering that the Covid response achieved or might reasonably have achieved. In Section 2.2, we argue that historical data on life expectancy reductions precipitated by large economic shocks, make it unclear whether the measures that were adopted will end up being net life savers, or achieve reasonably close to minimum achievable total life loss. Another indication of the reality of the trade-offs is that any reasonable estimate of the cost per life year saved from Covid by nonpharmaceutical interventions substantially exceeds the £30K per life year threshold usually applied by NICE (the UK National Institute for Health and Care Excellence) when approving introduction of a pharmaceutical intervention. For example, taking the half a million potential Covid deaths initially predicted for the UK under minimal mitigation (Ferguson et al., 2020) less the recorded Covid deaths to date would suggest around 300 thousand lives potentially saved. Given approximately one decade of life lost on average per victim (e.g. Hanlon et al., 2020), this corresponds to 3 million life years saved (16 life days per head). Taking this as the upper bound on life years saved, and a conservative £$10^{12}$ of extra borrowing plus lost economic activity as the cost of the interventions, gives a cost per life year saved over 10 times the NICE threshold. Given that health spending is necessarily finite, this comparison suggests a trade-off between life years saved from Covid vs. life years saved from other diseases that may not be straightforward to justify.

If one accepts the existence of significant trade-offs, the reality of the collateral damage and its nonnegligible size in relation to the benefits of the Covid measures, then in addition to identifying the many things that went well in the Covid response, it is important to discuss what went badly. Without such a dialogue, we reduce the chance of doing better next time. This paper is about openly discussing statistical aspects of some of what did not go so well. Data and code used in the paper are supplied in an online supplementary material R package.

## 2 The presentation of risk

> …a substantial number of people still do not feel sufficiently personally threatened; it could be that they are reassured by the low death rate in their demographic group…the perceived level of personal threat needs to be increased among those who are complacent, using hard hitting emotional messaging.

This unusual approach, of apparently intentionally distorting the presentation of medical risk in the service of a public health goal, is extracted from the 22 March 2020 recommendations of the UK government advisory Scientific Pandemic Influenza Group on Behaviour (SPI-B). The actual risk profile underlying the complacency is shown in Figure 1. The ethics of distorting risk perception in this way are open to question, particularly if done in the interests of promoting a 'greater good', subjected to detailed quantification of the short-term benefits, but not of the long-term disbenefits. One of the milder examples of the approach was a widely displayed government poster picturing a healthy woman in her mid-twenties in a mask with the slogan 'I wear this to protect you. Please wear yours to protect me'. The framing in terms of reciprocal risk implied either large overstatement of the risk to the person pictured, or a failure to consider her Covid risk in relation to her baseline risk. For example, the current best estimate for the return time of a super-volcanic eruption of the civilization-ending magnitude that city dwellers are unlikely to survive is 17 thousand years (Rougier et al., 2018). Even only considering the 2 years of the pandemic, this is likely larger than the Covid risk to the woman pictured.

[2] To quote former Bank of England chair Mervyn King (BBC, 23 October 2022): 'during Covid, when the economy was actually contracting because of lockdown, central banks decided it was a good time to print a lot of money … That led to inflation. We had too much money chasing too few goods, and the result was inflation. That was predictable. It was predicted, and it happened'.

[3] Close to 7% and increasing steeply before exacerbation by the Ukraine war.

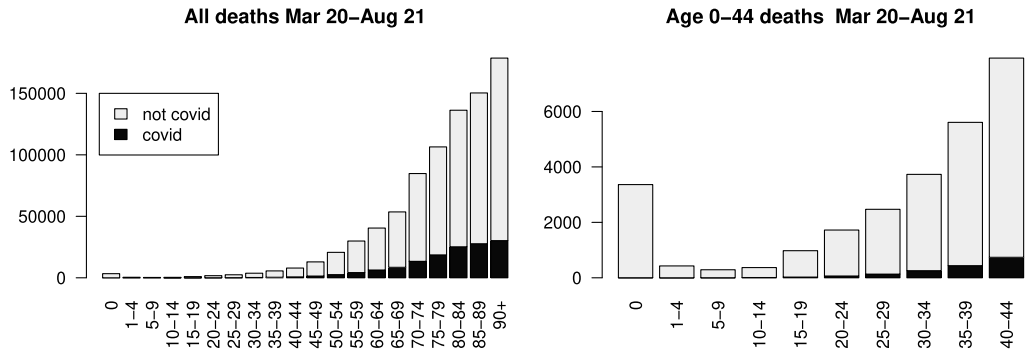



**Figure 1.** All UK deaths from March 2020 to August 2021 by age band. Covid deaths are shown in black. Data are from the Office for National Statistics. The right panel simply enlarges the plot for the under 45 s, where the risks are otherwise too low to be visible.

Figure 1 is based on retrospective data, and one moderately frequent argument is that at the start of the UK epidemic, the risk factors were not known. This is, at best, only partially true. Initially, the Chinese authorities were more open in sharing clinical data (e.g. Huang et al., 2020; Wang et al., 2020; Wu et al., 2020; Zhou et al., 2020, online publication: Jan 30, Feb 7, Mar 9, Mar 9) than they later were when it came to information pertaining to Covid origins. Statistical uncertainty notwithstanding, these studies provided a good deal of information on co-morbidity risk factors. Combined with the Diamond Princess cruise ship outbreak data, the studies also gave reasonably solid data on risk with age profiles (e.g. Verity et al., 2020; Wood et al., 2020, from March and May 2020).

The remainder of this section discusses how Covid risks were assessed and presented beyond the start of the first wave, as well as how authorities typically failed to present the health risks associated with the response.

## 2.1 The risk from long Covid

From mid-2020, the narrative around risk in the young and healthy began to shift towards the need to avoid long Covid, but again the presentation of the risk was questionable, and again hard-hitting emotional messaging was used, rather than presenting the actual risk. For example, in October 2021 the Department of Health and Social Care released a film on the dangers of long Covid, the press release including this warning about the risks after mild illness:

> Tom, 32, who features in the film[4] says: 'Do not make the mistake of thinking that being young or being fit is going to stop COVID from having a long-term impact on your health'.

As expected for any serious pneumonia (e.g. Herridge, 2011; Hopkins et al., 1999), hospitalized patients clearly suffered long-term effects (Ayoubkhani, Khunti, et al., 2021, for example[5]), and a novel virus was always likely to result in an increase in people suffering longer-term postviral complications (e.g. Appelman et al., 2024). However, beyond anosmia (loss or change in the sense of smell), the evidence for exceptional risk from Covid-specific sequelae after mild illness came from studies with substantial statistical problems, and it is difficult to view the level of concern about such sequelae as reflecting normal evidence-based medicine. One problem is the tendency to use catch-all definitions such as the NICE (2022) definition of *Post-COVID-19 Syndrome:*

> Signs and symptoms that develop during or after an infection consistent with COVID-19, continue for more than 12 weeks and are not explained by an alternative diagnosis.

[4] https://www.gov.uk/government/news/health-secretary-warns-of-long-term-effects-of-covid-19-as-new-film-released and https://www.youtube.com/watch?v=ulJSEo2fWvA. Tom is shown clearly face on, and there is no indication that his name has been changed or that he is being represented by an actor. He is also shown as having a Reading Half Marathon 2019 medal, but there are photos online of all the 4 Toms who completed that year, and none appear to be him (ditto Thomas or Tomasz).

[5] Although with a general population control group this does not provide evidence for a Covid-specific syndrome.

Obviously, such a definition invites *post hoc ergo propter hoc*. Almost any event could be substituted for *infection consistent with COVID-19* and a substantial number of cases of the associated syndrome would be found. High prevalence estimates were almost always based on a wide range of self-reported symptoms, often in nonrepresentative samples, with no control group, so that a meta-analysis of 174 studies produced up to January 2022 (O'Mahoney et al., 2023) reported at least 45% of Covid cases producing ongoing symptoms after 4 months. Amin-Chowdhury and Ladhani (2021), Haslam and Prasad (2023), and Høeg et al. (2023) all highlight the lack of control groups in most of these early studies, the vague and variable case definitions, and the use of samples that were not representative of the general population. As an illustration of the problems, an ONS study (Ayoubkhani, Pawelek, et al., 2021) found prevalence of long Covid symptoms at 5% in confirmed Covid cases vs. 3.4% in matched controls, suggesting a 95% CI for Covid associated symptom prevalence of (1.0, 2.2)%. The same study found *self-reported* long Covid prevalence of 11.7%. Nevertheless, the ONS continued to publish a survey of self-reported long Covid until March 2023 when a prevalence of about 2.9% was reported (Rea et al., 2023).

A further difficulty, even with carefully designed studies, is relatively low response rates and the associated serious risk of participation bias. For example, of the approximately 800,000 REACT study participants invited to take part in the REACT Long Covid study 276,840 responded (Atchison et al., 2023): there is an obvious danger that those with ongoing symptoms are more likely to participate. Similarly, a large-scale Scottish study (Hastie et al., 2022) had response rates of 15% for the controls and 20% for the cases: it only takes a proportion of the extra 5% in the case group to have participated *because* of symptoms, to produce the differential symptom rates observed between cases and controls, even if the real rates are identical. Prospective studies should, in principle, be more reliable. The prospective cohort case–control study of Ballering et al. (2022) found 381/1,782 Covid cases with at least one ongoing symptom at 90–150 days postinfection, vs. 361/4,130 for a control cohort matched by 'event' time, binarized age and sex, suggesting a Covid attributable rate of 12.7%, in a 6:4 female skewed cohort with age distribution over-concentrated around mean 54. But dropout was high. The initial Covid cohort size was 4,231, so there is considerable scope for differential dropout between those with and without ongoing post-Covid symptoms to have skewed the rates.[6] Perhaps more concerning is that, except for ageusia/anosmia, the 'core symptoms' employed in the paper are all from the somatization sub-scale of the Symptom Checklist-90 questionnaire, used to assess psychological problems which may be expressed somatically (see. e.g. Holi et al., 1998; Van Driel et al., 2018). To be sure that these are physical sequelae to Covid itself requires subject blinding to their Covid status. In any case, the initial cohort included 142 hospitalized cases (final number not given) and the final cohort contained 158 cases with ageusia/anosmia. Hence, from the information presented, it is not possible to work out rates for Covid attributable symptoms excluding ageusia/anosmia in nonhospitalized patients, or even to rule out these being zero.

In fact, for nonhospitalized cases, carefully designed studies often found low, or even no, differences in the frequency of persistent symptoms in cases and controls, except for anosmia. For example, a systematic review and meta-analysis of 22 studies of children and young people (Behnood et al., 2022) reported that the frequency of most reported persistent symptoms was similar in SARS-CoV-2 positive cases and controls. A study in Norway (Selvakumar et al., 2023) using the WHO definition of 'post-COVID-19 condition' found that prevalence at 6 months was similar in test-positive nonhospitalized cases and test-negative controls. In a large population-based French cohort for ages 18–69 (Matta et al., 2022), self-reported Covid-19 infection was associated with persistence of multiple physical symptoms 10–12 months after the first wave, whereas laboratory-confirmed Covid-19 infection was associated only with anosmia. The authors suggest that persistent symptoms 'may be associated more with the belief in having experienced COVID-19 infection than with actually being infected with the SARS-CoV-2 virus'. In a study comparing ongoing symptoms post-Covid and post-Influenza, Brown et al. (2023) found no difference in symptom rates between the two cohorts. None of this is to deny the existence of real

[6] E.g. if the symptom rate was 9% in cases and controls, but almost all those with ongoing symptoms post-Covid stayed in, then the results would be very close to those obtained. Loss of interest dropout in the fully recovered is also not implausible.

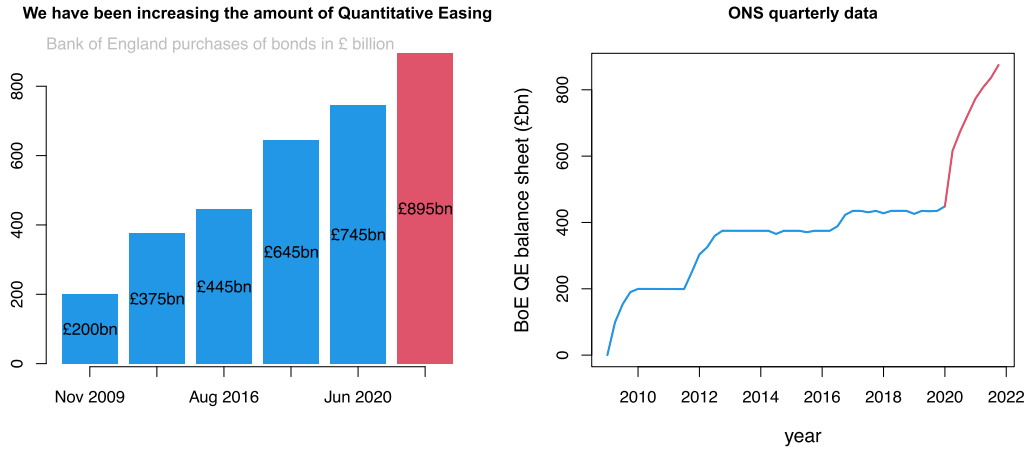



**Figure 2.** Left: The Bank of England quantitative easing (QE) programme as presented on the Bank of England website in 2021 (redrawn, but identical in all relevant respects). Right: the same QE programme using a continuous linear time scale (ONS data), with the Covid part shown in red.

sequelae to Covid infection, but rather to suggest that the long Covid evidence base was of insufficient statistical quality to form part of the justification for the continuation of severe societal restrictions, or of risk distorting public messaging. Looking forward, Høeg et al. (2023) make recommendations for improving the situation, to both avoid irresponsible exaggeration of risk after mild infection, and to better serve those suffering long-term problems genuinely caused or triggered by Covid infection.

The referees also suggested discussing the economic effect of long Covid, in particular, on UK employment levels. The background is that people of working age not in employment as a result of long-term illness has increased by around 700,000 since the start of the pandemic, an increase of around a third of the prepandemic level. ONS (2023) and the associated data set reports estimated increases in health-related inactivity by primary cause, stating that long Covid is classified under 'Other health problems or disabilities', which increased by 156,000 between 2019 and 2023. This is some 80,000 more than would be expected from the general increase, suggesting an upper bound for long Covid's impact on long-term inactivity. An alternative estimate comes from ONS (2022), which estimated a 0.5–3.4% increase in economic inactivity among those self-reporting long Covid suggesting 10–70 thousand people with long Covid related inactivity if we take the contemporaneous ONS estimate of 2 million people with self-reported long Covid as all of working age (an overestimate). For comparison, 593 thousand report musculoskeletal or connective tissue problems and 645 thousand depression or other mental disorders as their primary causes of economic inactivity. Of course, those economically inactive from long Covid represent only a proportion of those substantially impacted, but this is equally true of the other medical causes of inactivity.

## 2.2 The risk of life loss from economic shock

While *some* of the risks from Covid were being exaggerated, the risks from the measures taken were downplayed. The left panel of Figure 2 is faithfully redrawn from the Bank of England website.[7] Its rescaling of the time axis creates the impression that the quantitative easing (QE) programme (funded by money creation) had expanded steadily over time, with the Covid part simply following an existing trend. The right panel, using a continuous linear time axis to plot the same data, creates a rather different impression.

Discussions that involved the size of the QE programme were often framed in terms of 'saving life vs. saving the economy'. Such framing relies on not considering another set of data: that relating to the health effects of economic deprivation and inequality. A major evidence-based review of

[7] www.bankofengland.co.uk/monetary-policy/quantitative-easing—graphic removed in late 2022.

these links had been published early in 2020, before the pandemic (Marmot et al., 2020). That review updated an equally substantial 2010 report on the links between economic deprivation and health inequality. The 2020 report includes a forensic, data-based investigation of how health outcomes and life expectancy for the more disadvantaged were worsened by the exacerbation of economic deprivation following the financial crisis of 2008 and subsequent government response to it (the latter democratically sanctioned by the 2010 election[8]).

The left panel of Figure 3 shows how life expectancy changes with population-weighted decile of area deprivation in England (from Marmot et al., 2020, Figure 2.3). Taking the upper 10% as representing the current potentially achievable life span, the plotted data correspond to a, presumably avoidable, life loss of 3.2 years per capita for women and 3.95 years for men. If the mean life expectancy of the upper half is taken as the achievable figure, then the loss reduces to about 1.8 and 2.5 years, respectively. Scaling up to the current UK population, this is a potentially avoidable life loss of 140–290 million years. Whether one is politically inclined to view this life loss as avoidable or not, any small percentage increase to it, as a result of the economic disruption caused by the Covid response, obviously risks life loss on a scale comparable to that from Covid itself.

How likely is such an increase in life loss? Internationally, there are clear examples of linkage between economic problems, deprivation, and lifespan reduction (Case & Deaton, 2017; Ciment, 1999; Ruminska-Zimny, 1997). Also, Stuckler et al. (2010) point out an association between welfare spending and mortality strong enough, if it contains a causal component, to imply that any crisis either directly reducing spending or preventing increased spending would result in substantial avoidable life loss.[9] Recent historical data for the UK also give some indication. Marmot et al. (2020) argue that the response to the 2008 financial crisis is implicated in a reduction in the trend for increasing life expectancy equivalent to around 1 year life loss per capita. If the larger financial shock from the Covid response produces knock-on effects anything close to this, then the measures will have cost far more life than they saved. However, this figure is based on assuming that an apparent linear trend in life expectancy before 2010 would simply have continued afterwards, in the absence of the 2008 crisis and response. What actually happened to life expectancy is being compared to the extrapolation of a purely statistical model.

A much more conservative approach is to base all comparisons on what actually happened to life expectancies: to take the post-2008 life expectancy trends among the more prosperous as the measure of what could reasonably be expected for those not substantially impacted by economic effects, rather than straight line extrapolations. Then, we can ask how the economic deprivation life expectancy gap in fact changed after the 2008 crisis. Treating the least deprived 10% as the 'more prosperous' group, the right-hand side panel of Figure 3 shows the results of such a difference in differences approach. The corresponding increase in avoidable life loss is about 7–9 weeks per capita, or 9–12 million life years for the whole UK population, and the evidence presented in Marmot et al. (2020) makes it difficult to discount the knock-on effects of the crisis having had a very direct role in this. It is unclear that there are any solid reasons to expect the Covid response economic deprivation-related life loss effects to be smaller.

One objection to this difference in differences approach is that for the most part life expectancy improved somewhat across the deprivation scale after 2008, but more slowly for the more deprived. The argument is, therefore, that the loss is not 'real'. But such a position has perverse consequences. Even post-2008, average UK life expectancies grew by around 18 days per year. Is any per capita life loss in a year that is less than this not real? The direct loss from Covid over 2020 amounted to about 6 days per capita, and it would be absurd to say that there was no real loss of life because this is less than the expected 18-day life expectancy increase for the year. Similarly, deprivation has caused real loss of life, even though the average life expectancy has increased.

A final objection is that the exacerbation of economic deprivation after a financial shock is not inevitable, but is a political choice: hence these data are irrelevant in considering whether to impose the shock, and are only relevant to how its consequences are later dealt with. This seems

[8] See, e.g. p. viii 2010 Conservative party manifesto, p. 14 2010 Liberal Democrat Manifesto.

[9] They found a 1.2% reduction in all-cause mortality for each 100 USD per capita increase in welfare spending. The increase of around £1,000 per head in annual government debt servicing costs since 2020 is obviously unavailable for increased welfare.

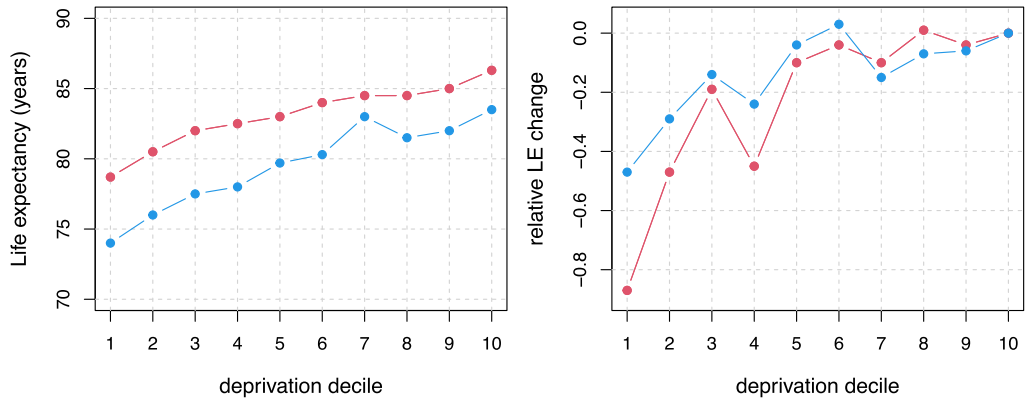



**Figure 3.** Left: Life expectancy by (population-weighted) area deprivation decile for women (red) and men (blue). Right: The change in years of life expectancy relative to the least deprived between 2010–2012 and 2016–2018 for women (red) and men (blue). Data from Figures 2.3 and 2.5 of Marmot et al. (2020), which were based on ONS and PHE data.

to us to be a counsel of perfection, firstly because there are inevitably some economic constraints on what is politically possible (money spent on debt servicing is not available for welfare, for example), and secondly, because the historical data probably offer a better guide to what is politically likely than utopian considerations of what political action could ideally achieve. Whatever one's views about the policies adopted in the UK after 2008, they are what the electorate explicitly voted for in 2010. In other words, there is a real risk that the electorate may not vote for the 'best' policies after an economic shock, or that no political party offers such policies. Should such a real-world risk be ignored?

This is, of course, not to argue that a substantial risk equates to certainty. After the profound existential trauma and economic dislocation of the Second World War the electorate did vote for a substantial enhancement of the welfare state. However, this expansion cost less than 3% of GDP when annual growth was 3–4%, and substantial US aid was available, so that, despite the economic situation, government debt began falling sharply immediately after the war ended (see Crafts, 2023, for more detail). The contrast with current circumstances implies a need for circumspection in reading this precedent. The postwar reforms also point, perhaps, to the limits of what even very enlightened policy can achieve, postshock. Although the detailed mechanisms are, of course, complex, the ratio of standardized mortality rates in social classes V to I had actually increased in 1949–1953 relative to the prewar figure (Pamuk, 1985). Similarly, the government commissioned 1980 Black report[10] found that health inequalities had increased since 1948 (Gray, 1982). The fact that Black's recommendations for improvement were not implemented, while many of the later Acheson (1998) report were, may at least partly reflect the difference between what was politically possible in the recession of the early 1980s vs. the strong economy of 1998, although differences in government (and electorate) political philosophy were also significant.

The economic deprivation figures and the very approximate 3 million life years potentially losable to Covid were available by mid-March 2020. Clearly, the path from economic shock to substantial life loss is uncertain and very difficult to credibly model. This does not mean that there is no evidence that the effect exists. It only means that it comes with a range of possible loss of life and associated uncertainty, which is the very basis of the definition of risk. In short, the available data in early 2020 indicated that a large economic shock would come with a substantial risk of downstream loss of life.

## 2.3 The meaning of life expectancy

The life expectancies discussed above broadly have the interpretation that if things stay much as they have been over the preceding few years, then this is how long we can expect to live, provided

[10] https://sochealth.co.uk/national-health-service/public-health-and-wellbeing/poverty-and-inequality/the-black-report-1980.

nothing changes drastically. The changes in life expectancy discussed by Marmot et al. (2020) are of this slowly varying nature, which is why it is reasonable to use them in calculations involving potentially avoidable life-year loss. However, during the pandemic, media reports often stated that Covid had caused a life expectancy drop of around 1 year, while almost always omitting the qualifier that to interpret this as indicating that the average UK resident's expected lifespan had been shortened by a year would be entirely false. The 1 year drop is what would happen if there was a new Covid epidemic, causing comparable life loss, every year from 2020 onwards. Unqualified statements such as

> Americans are now expected to live an average of 77.3 years, down from 78.8 years in 2019[1]

or later and hence less dramatically,

> A boy born between 2018 and 2020 is expected to live until he is 79, down from 79.2 for the period of 2015–17,[12]

not only omit the caveat but contradict it, explicitly interpreting *period* as *cohort* life expectancy.

The media reports were based on scientific papers that perhaps assumed that their readership understood the caveat, without the need to explicitly state it. A case in point is Islam et al. (2021), who reported both reductions in 'life expectancy' and life years lost for 37 countries for 2020. For example, they report a 2 year life expectancy drop for Bulgarian men, who were also estimated to have lost 7,260 life years to Covid per 100 thousand population. The latter figure corresponds to an average life loss of 4 weeks per head. The equivalent figures for the UK in 2020 were a life expectancy drop of about 1 year and a life loss of about 6 days per head.[13] When assessing risk, the difference between risks that would shorten your expected lifespan by 1 year vs. 6 days is quite substantial. Presenting a figure in a way likely to lead to a 60-fold overestimation of actual risk seems unlikely to promote a proportionate response to the risk.

## 2.4 Excess deaths

A useful way of calibrating Covid risk retrospectively is provided by computing excess deaths: the excess of actual deaths over what might reasonably be expected given the situation in previous years. But here too there is scope for statistical confusion. Excess deaths are often computed by looking at deaths relative to seasonal averages over a number of years preceding the period of interest, with various adjustments made to account for trends in mortality over time. That adjustments are needed is clear if the postwar baby boomers are considered. For the UK, the year group conceived immediately after the war is 31% larger than the year group from the previous year (see Figure 5). These people were approaching 75 at the start of 2020. Failure to take account of this demographic cliff edge advancing into the age group at which mortality rises sharply with age is bound to lead to inflation of apparent excess death rates.

We argue that the simplest approach to excess deaths is to take life tables computed from the mortality data over a reference period of the years immediately preceding the period of interest, along with the population's age structure at the start of the period of interest, and to simply iterate the ageing and death processes forward in time. Applying the same process from the population structure at the start of the reference period offers the sanity check that the total deaths over the reference period should match between data and predictions. See Appendix A for details.

Figure 4 does this for the UK, with 2017–2019 as the reference period, and all data obtained from the Office for National Statistics. Ageing and death are applied weekly to weekly age cohorts. The yearly cycle is obtained by fitting a generalized additive model to the reference period data with a smooth for week, and a cyclic smooth for week of the year. From this, a multiplier relating

[11] https://www.cnbc.com/2021/07/21/life-expectancy-in-the-us-declined-in-2020-especially-among-people-of-color-.html.

[12] https://www.theguardian.com/society/2021/sep/27/covid-has-wiped-out-years-of-progress-on-life-expectancy-finds-study.

[13] The current total from the start of the pandemic stands at about 12 days per head, under the assumption of a decade of life loss per victim and given the UK government figure of 226 thousand deaths with Covid.

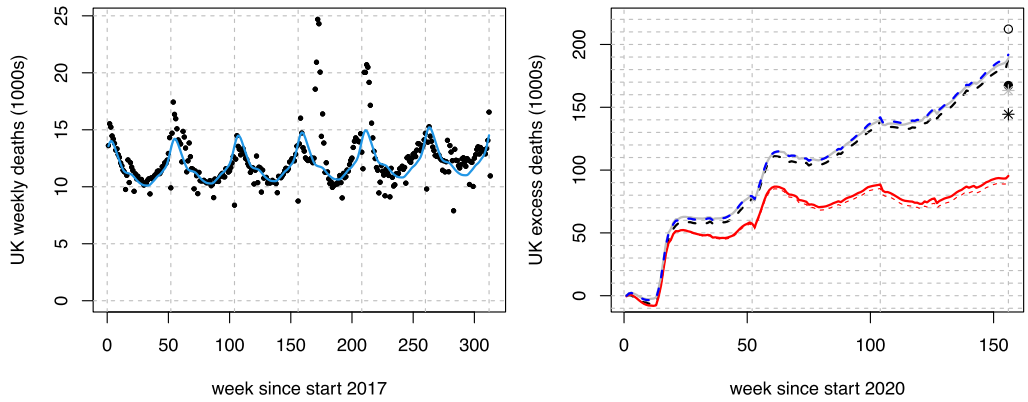



**Figure 4.** Left: Weekly UK deaths (black dots), and weekly deaths predicted using 2017–2019 life tables (blue). The predictions start from the age structure of the population at the start of 2017 for the first 3 years, and the age structure at the start of 2020, thereafter. The cumulative observed and predicted deaths match to within 0.0025% (50 deaths) over 2017–2019. Right: red solid is the cumulative excess deaths from the start of 2020 iterating from population by age data for the start of 2020; red dashed is the equivalent iterating 6 years from population by age data for the start of 2017; black dashed is for the conventional method based on 2017–2019 data; grey is the same but using life table iterated weekly deaths in place of raw deaths; blue dash is the life table-based approach with ageing turned off, starting from the estimated population by age in mid-2018. The open circle is the government figure for deaths with Covid, the black disc is the ONS pandemic excess death figure, the star is the PHE pandemic excess death figure for England (grey scaled up prorata to the UK). We argue that the red curve is the most reasonable, because it directly accounts for the effects of population ageing on the expected deaths.

the weekly death rate to the annual death rate can be computed, so that weekly death rates can be used in the iteration of the demography. Note that the total number of deaths predicted by this iteration over 2017–2019 matches the observed to within 50 deaths (0.0026%), when the iteration is started from the age-structured population at the start of 2017. The figure of around 95 thousand total excess deaths from the solid red curve in the right panel of Figure 4 is lower than the figure of 167,356 given by the ONS as the excess death figure from March 2020 until the end of 2022, or the PHE figure for England of 144,446 (equivalent to around 163 thousand if crudely scaled up to the UK). The ONS figure is based on simply comparing weekly deaths to the average for that week of the year for the 5 years preceding 2020, with PHE similar but based on statistically modelled deaths. These figures obviously neglect the consequences of baby boomer and general population ageing that lifetable iteration includes. That this ageing effect is indeed large can be confirmed by a very simple check, which is shown graphically in Figure 5. The left-hand side plot compares the population in each age class from age 50 onwards for 2017 and 2020, according to ONS. For each age group, the growth in population from 2017 to 2020 can be multiplied by the annual death rate for that age group (from ONS life tables), and then summed to get the expected change in a number of deaths per year that ageing has caused. The total is about 30 thousand expected extra deaths per year and the right-hand side plot, of cumulative changes in death with age, illustrates how the change is accumulated across the age groups. Note that migration makes a negligible contribution to these figures, given that only some 7% of migrants are over 50 and 'negligible' numbers over 70 (Home Office, 2021).

Since the ONS and PHE figures are based on a 5-year time window, the right panel of Figure 4 also shows the results of applying the standard weekly death rate averaging method using 2017–2019 as the reference period. The total excess deaths predicted by this method actually exceed the ONS and PHE figures, which is unsurprising given that 2015 was a relatively high death year and 2019 relatively low. To further emphasize the dependence of the standard method on neglecting ageing, it is also possible to modify the iterated lifetable approach by turning off ageing (while also having deaths not deplete the age groups, so that the overall population does not decline). Applying this, obviously deficient, process, starting from the mid-2018 age structure, gives a close match to the current standard methods.

Given the size of the ageing effects, it is difficult to see that the iterated lifetable approach does not give the more reasonable expected number of deaths, relative to the current standard methods.

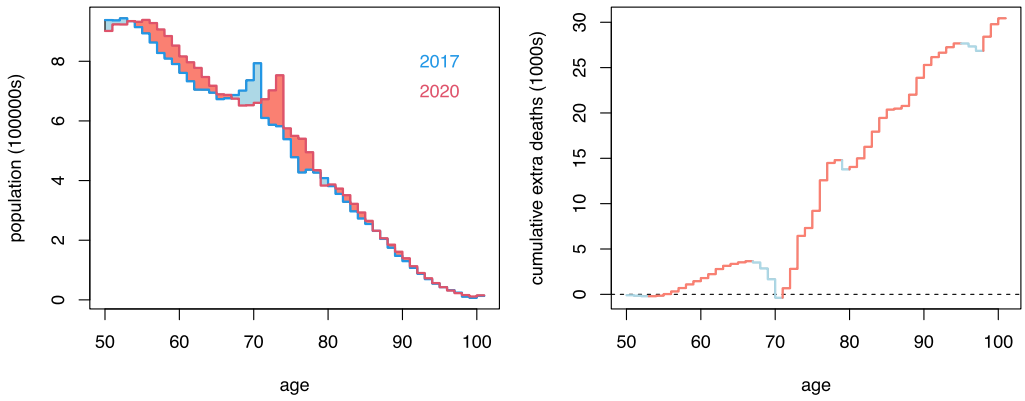



**Figure 5.** The effects of population ageing on deaths. Left: Blue shows the UK population by the 1 year age band in 2017, and red shows the equivalent in 2020. The difference between the two ($\Delta_i$ for age class $i$) is shaded blue when the 2020 population is less than the 2017 population and salmon when the 2020 population is larger. Right: The population difference, $\Delta_i$, in each year class between 2020 and 2017 is multiplied by the annual probability of death in that class, $m_i$. The cumulative sum of the resulting expected extra deaths per year (i.e. the expected extra deaths among those from 50 up to age $i$, $d_{\leq i} = \sum_{j=50}^{i} m_j \Delta_j$) is then plotted against age class, with colour coding corresponding to the left plot. The 3 years of population ageing leads to 30 thousand extra expected deaths, a large effect that can obviously not be captured by the traditional approaches to computing the expected deaths for excess death calculations. Population data and annual per capita mortality rates at age are from the ONS.

The interpretation of the resulting expected deaths is also particularly clear: *if age-specific death rates remained unchanged from the reference period, then this is the number of deaths that would occur*.[14]

The cumulative excess deaths, shown in red in the right panel of Figure 4, are much lower than the total deaths recorded with Covid (212,247 with Covid mentioned on the death certificate by the end of 2022, according to the UK government's data dashboard). There are a number of mechanisms that are likely to account for this. An obvious one is the fact that only some 17 thousand people had only Covid and nothing else recorded on their death certificate. When Covid is only one factor among several in a death, it is statistically naive to expect it to contribute a whole extra death in the excess figures (given that some of the other factors are risks contributing to what is expected without Covid). Put slightly more technically, since dying with Covid and dying with other comorbidities are not independent events, Covid mortality events do not simply add to the mortality caused by the other comorbidities. Related to this are what epidemiologists refer to as 'harvesting' effects: where an epidemic pathogen brings forward the deaths of some very frail people by only a few weeks or months. Over a period of 3 years, many such people will not appear as excess deaths at all, since their death has only been moved *within* the time period considered.

For near real-time monitoring of excess deaths, it has been argued (e.g. Holleyman et al., 2023) that expected deaths should be corrected for harvesting effects—that someone who was expected to die shortly from other causes, but succumbs to Covid earlier, should then be removed from the later expected deaths. The approach obviously has some philosophical difficulties, since dying of any cause means that you do not die later of some other cause: hence a decision has to be made to treat only some causes of death in this way. We, in effect, *define* Covid deaths as excess.[15] A second problem with using the approach to compute excess deaths over an extended period is the change in interpretation of the statistic. Conventionally, excess deaths in a period are the excess of observed deaths, $D$, above the number expected over the period given the mortality data up to the start of the period, $E$. Consider the case in which some horrible event led to 10,000 unexpected deaths of otherwise healthy 5-year-olds in a year. Both the Holleyman method and the conventional method count these as 10,000 excess deaths in the year. Now consider the case in which $D$ is 10,000 less than

[14] Note that, at about 140 per 100 thousand, our figures are still higher than the 120 per 100 thousand estimated excess deaths over the two pandemic A/H3N2 influenza seasons starting in 1968 (Viboud et al., 2005).

[15] One problem with this is that many lives of the terminally frail are eventually ended by opportunistic infections: focusing on the particular infectious agent is not always meaningful in such cases.

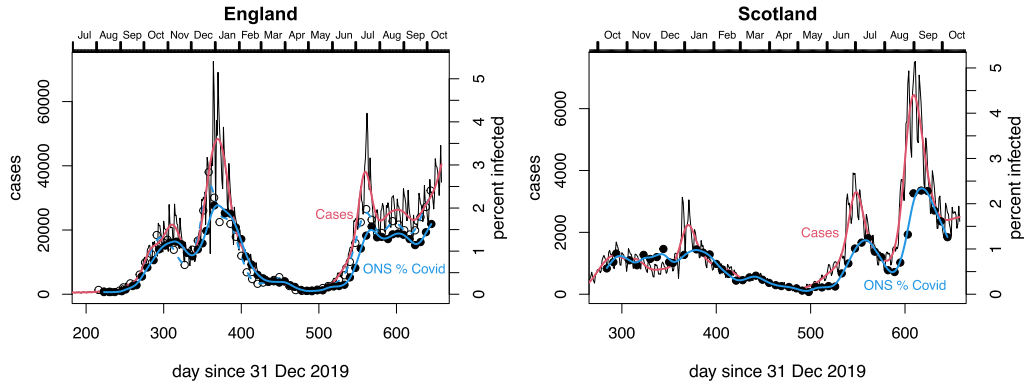



**Figure 6.** Comparison of official daily Covid cases data (black lines, red smooth) and ONS prevalence measurements (black dots, blue smooth) for England on the left and Scotland on the right. In both cases, the data are scaled to match over the 50 days from the midpoint day 430. If case data reliably measured prevalence, then the smooth curves should be uniformly close to each other. For England, the ONS incidence estimates (scaled) are also shown, as open circles smoothed by a dashed blue curve.

*E*, but 20,000 people die one day earlier than expected from the cause selected under the Holleyman method. The Holleyman method again produces 10,000 excess deaths, as under the first scenario, although in total, we saw 10,000 fewer deaths than were expected to occur over the year. A further problem is what to do about the situation in which occurrences such as a low respiratory pathogen season (or lockdown) displaces some deaths to *later* than they would have been otherwise expected to occur. If we correct for people who die earlier than expected due to an unusually high risk from pathogens, should we not also correct for those who die later than expected due to an unusually low risk from (other) pathogens? Focusing on excess life-year loss would avoid *some* of these problems, but it is not easy, given available data, without strong modelling assumptions.

An objection to the explanation that, in many cases, Covid may have brought forward deaths by 'only' a few months is that studies looking at life loss per Covid victim suggest figures of around a decade on average (e.g. Hanlon et al., 2020), which would imply a more limited role for harvesting. However, such life loss studies tend to suffer from the problem of having to treat comorbidities as simple categories (often binary), with limited possibility for incorporating comorbidity severity, leading to likely inflation of the estimated life loss per victim. For example, suppose that each victim with congestive heart disease as a comorbidity is assigned the average life expectancy of someone with congestive heart disease. Then, we will overestimate life loss to Covid, if, in reality, it is those with more severe disease, and consequently shorter Covid free life expectancy, who are most likely to succumb to Covid.

## 3 Covid cases and other media distortions

A statistically troubling feature of the media and government presentation of the state of the Covid pandemic was the preference given to 'case' data, even after the ONS had started directly measuring prevalence using sampling, from mid-2020 onwards. Case data were discussed as if they were proportional to prevalence, although it is unclear what population they sample and how they relate to prevalence. The data measure people testing positive among those who were tested: largely those who chose to be tested, or were advised to by test and trace, and could obtain a test. How this number relates to prevalence at any given time is somewhat obscure: some relationship is to be expected, but it seems unlikely to be constant, in large part because of the way testing behaviour was likely to change over time, for example, in relation to both perceived risk and available testing capacity. Figure 6 shows the correspondence between prevalence measurements and case data over time, scaling the data to match over a time interval in the middle of the period. The focus on case data creates a clear danger of over-reacting when prevalence is increasing, as numbers of infections appear to climb more steeply than is actually the case.

It can also be argued that case data should be proportional to incidence, not prevalence (although, for a disease of 2 weeks or so duration, the distinction is a rather fine one, given the

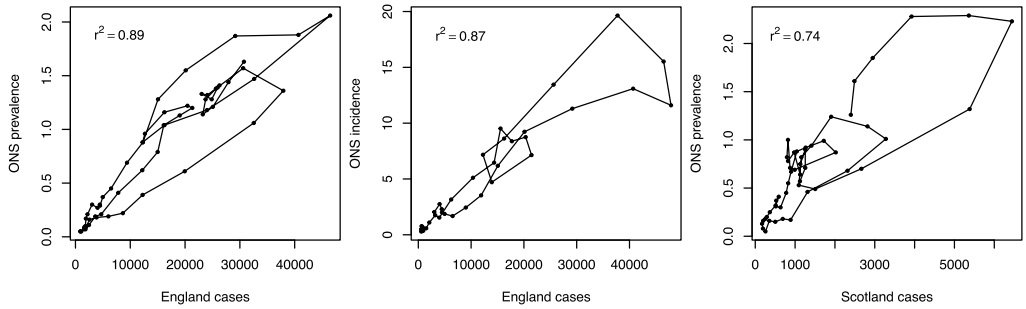



**Figure 7.** Left: ONS prevalence plotted against cases (weekly cycle smoothed out) for England, with points joined in time order. Middle: similar, with ONS incidence on the vertical axis. Right: as left but for Scotland. The shifting nature of the relationship between cases and prevalence or incidence is clear.

difficulty identifying the population being sampled). This introduces the additional problem that there is an epidemiologically significant and variable delay from infection (the event relevant for incidence) to detection as a case. The left panel of Figure 6 also shows the ONS surveillance survey reconstructions of incidence for England, again scaled to match cases over 50 days from the data-period midpoint. Despite the generally poor correspondence, there are periods on the upswing of a wave when the cases match scaled incidence quite well, before the cases overshoot. This match probably results from the over-acceleration of case detection compensating for the delay from infection to case detection—it is unlikely that anyone would argue that it is prudent to rely on such a fortuitous cancellation of biases.

These problems have not prevented some authors from continuing to argue that cases represent a good proxy for actual prevalence or incidence. For example, Brainard et al. (2023) make this case on the basis of the correlation coefficients between cases and ONS estimates. Figure 7 illustrates the substantial systematic drifts in calibration that lie behind the correlations.

A curious argument was sometimes advanced that the case data were reliable for assessing trends in infection rates and the pathogen reproductive number, $R$, *if looked at over a short enough time window*. This seems equivalent to falsely asserting that $\lim_{\Delta\to 0}\Delta^{-1}\{f(t+\Delta)+g(t+\Delta)-f(t)-g(t)\} = \mathrm{d}f/\mathrm{d}t$, where $g(t)$ is some function deemed inconvenient. That concern over such matters is more than nit-picking is illustrated, for example, by Liu et al. (2021), who attempted to model the effects of various nonpharmaceutical interventions (NPIs) on $R$, which was estimated from case data. Their Table 5 summarizes the effects of the 13 NPIs considered. All apparently reduced $R$, apart from trace and trace, which apparently increased it. It seems improbable that trace and trace actually increased transmission, and substantially more likely that it accelerated the finding of cases leading to inevitable upward bias in the case-based $R$ estimate.

Poorly founded opinions are to social media what sand is to a beach, so we will not comment on what appeared there. But it is reasonable to expect better from traditional media with a reputation for journalistic integrity, fact-checking and editorial control. This expectation was not met by, for example, the BBC's failure to put Covid deaths in the context of the average number of daily deaths from all causes, or by the Guardian having Nassim Taleb write about a historically rather moderate pandemic as if it were a 'black swan event'—a perspective unlikely to promote a balanced view of the actual risks. Another Guardian article from 27 April 2020[16] compared Covid to the black death, concluding that Covid was in some ways worse and

> …if we 'open up the economy' to help Tory grandees make money, there won't be much economy left once the second wave of infections has finally settled down, because all the Topshops will have to be razed to make space for graves.

The black death is estimated to have killed 30–60% of Europe's population irrespective of age. For Covid in the UK to be somehow equivalent, it would have had to cause a life loss of some 20

[16] https://www.theguardian.com/commentisfree/2020/apr/27/business-lockdown-johnson-tory-donors.

years per capita. The current figure actually stands at less than 2 weeks per capita. Even at the end of 2021 exaggeration was still common: for example, the *yahoo!news* headline 'Omicron: Germany records highest COVID daily death toll in nine months' was difficult to view as balanced at a point in time at which Germany had four Omicron cases, all very much alive. Most media commentators, of course, avoided such statistically nonsensical hyperbole, but the more moderate often still repeated that Covid was 'the worst pandemic for a century', an odd view at a time when LBGT rights and black lives matter were to the fore in public consciousness: the WHO estimates the AIDS death toll at 27–48 million, again with a high life-year loss burden per death.

In some ways, a more statistically concerning example was a 19 April 2020[17] article in the Guardian from two professors of biostatistics discussing the supposed difficulty of estimating prevalence:

> Arguably, the most important problem is the 'denominator'—what is the actual number of people who are infected by the virus? This is virtually impossible to determine, except perhaps in the unlikely scenario of real-time, continuous, population-wide testing.

The Guardian was not interested in printing a short letter correcting this statistically unusual perspective, by pointing out that randomized sampling could be used (as the ONS and REACT subsequently did).

That the media gets things wrong is perhaps why we have independent fact-checkers, but it is not clear that they, too, were not over-hasty in their judgements about which Covid narratives were 'correct', at least in part reflecting a lack of statistical knowledge. For example, in a lengthy article on evidence for lockdown efficacy fullfact.org asserted that the reason Wood (2020) had UK new infections per day peaking well before lockdown, while Flaxman et al. (2020) had surging growth in infections until lockdown (see Section 5) was that the former had assumed a much longer infection-to-death duration than the latter. That would be a compelling argument, were it not for the fact that the papers had used essentially the same infection-to-death distribution (Verity et al., 2020). fullfact.org were not interested in correcting this, nor some other incorrect statements about infection-to-death timings relating to apparently not understanding right truncation. They were also not interested in updating their article in light of the REACT-2 and ONS incidence reconstructions covered in Section 5.

This type of misleading and selective use of statistical evidence was not limited to the media. For example, in 2021, the official online Scottish government advice on face coverings stated that

> Scientific evidence and clinical and public health advice is clear that face coverings are an important part of stopping the spread of coronavirus.

and provided a link for the scientific evidence. This turned out to be a SPI-B/SAGE (Scientific Advisory Group for Emergencies) advice summary,[18] which cited two pieces of scientific evidence, apparently suggesting transmission reductions from mask-wearing of 6–15%, or up to 45%, respectively. The paper cited as evidence for the first figure was, in fact, an editorial (Cowling & Leung, 2020), which also pointed out that the paper cited for the 45% figure (Mitze et al., 2020) was flawed (the design appears unable to pick up the case in which mask-wearing is actually harmful, for example). The editorial's figure is quoting a properly conducted meta-analysis (Brainard et al., 2020) which actually concluded

> … wearing a mask may slightly reduce the odds of primary infection with [Influenza Like Illness] by around 6 to 15% […] This was low-quality evidence…

## 4 Epidemic dynamic models

Perhaps the most surprising feature of the epidemic models used to justify Covid policy was the omission of the fundamental role of person-to-person transmission rate heterogeneity investigated by Novozhilov (2008), and explicitly raised as a serious issue for Covid models in early 2020 by

[17] https://www.theguardian.com/commentisfree/2020/apr/19/coronavirus-deaths-data-uk.

[18] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/948607/s0995-mitigations-to-reduce-transmission-of-the-new-variant.pdf.

Gomes (eventually published as Gomes et al., 2022). The degree of variability between people in their susceptibility, connectivity and other determinants of transmission probability profoundly affects the size of epidemic—or of epidemic waves—predicted by the Susceptible Exposed Infectious Recovered (SEIR—exposed are infected but not yet infectious; recovered actually includes dead) type models that were used. The mechanism is simple: those individuals most susceptible to infection or most socially connected are preferentially removed from the susceptible population first, leading to a much more rapid reduction in infection rates than simple depletion of a susceptible population of clones would produce. Realistic levels of variability can easily halve the predicted epidemic (wave) size, and yet the models only accounted for the very modest heterogeneity in mean contact rates with age. It is possible that the early work on this topic was inaccessible, so we present the mathematical fundamentals of the mechanism here.

## 4.1 Person-to-person variability in SEIR models

First, let $\alpha$ be a parameter determining susceptibility to infection, which varies over the susceptible population, and let $s(\alpha, t)$ denote the susceptible population per unit $\alpha$ interval with parameter $\alpha$ at time $t$. Without loss of generality, we can scale the problem so that the initial population is 1, in which case $s(\alpha, 0)$ is the initial p.d.f. of $\alpha$. The standard SEIR model for this situation is,

$$\frac{\mathrm{d}s(\alpha, t)}{\mathrm{d}t} = -\alpha s(\alpha, t)I(t), \quad \frac{\mathrm{d}e(\alpha, t)}{\mathrm{d}t} = \alpha s(\alpha, t)I(t) - \delta e(\alpha, t), \quad \frac{\mathrm{d}i(\alpha, t)}{\mathrm{d}t} = \delta e(\alpha, t) - \gamma i(\alpha, t)$$

where $I(t) = \int i(\alpha, t)\,\mathrm{d}\alpha$. On integrating the first ordinary differential equation (ODE) we have $s(\alpha, t) = s(\alpha, 0)\exp(-\alpha q_t)$ where $q_t = \int_0^t I(t')dt'$. Since $q_t$ is monotonic in $t$, it is immediately clear how the epidemic progresses faster in subpopulations with higher $\alpha$. *In itself, this observation suggests that great care is needed in extrapolating to the whole population from those who become sick first.*

We can now obtain the total susceptible population at time $t$ by integrating out $\alpha$

$$S_t = \int s(\alpha, 0)\exp(-\alpha q_t)d\alpha = M(-q_t)$$

where $M$ is the moment generating function of the initial distribution of $\alpha$ (by definition). Now consider the time derivative of $S_t$,

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\int \alpha s(\alpha, t)d\alpha I(t) = -M'(-q_t)I(t) = -M'\{M^{-1}(S_t)\}I(t)$$

where $M^{-1}$ is the inverse function of $M$. So the SEIR dynamics are determined by three ODEs, without explicit dependence on $\alpha$.

Variability in contact rates can be modelled in a similar way. It is assumed that transmission depends on the product of $\alpha$ for the susceptible and $\alpha'$ for the infected. In this case,

$$\frac{\mathrm{d}s(\alpha, t)}{\mathrm{d}t} = -\int \alpha\alpha' s(\alpha, t)i(\alpha', t)d\alpha' = -\alpha\bar{\alpha}'_t s(\alpha, t)I(t)$$

where $\bar{\alpha}'_t = \int \alpha' i(\alpha', t)/I(t)d\alpha'$. Analytical progress now requires the approximation that the infectious state is short enough that the distribution of $\alpha$ in the infectious stage at $t$ is proportional to the distribution in those first becoming infected at $t$. That is $i(\alpha, t) = \alpha s(\alpha, t)$, so that $\bar{\alpha}'_t = \int \alpha^2 s(\alpha, t)\,\mathrm{d}\alpha / \int \alpha s(\alpha, t)\,\mathrm{d}\alpha$. If we now redefine $q_t = \int_0^t \bar{\alpha}'_{t'} I(t')d\alpha'$, then the maths follows through similarly to the variable susceptibility case, with the addition that $\bar{\alpha}'_t = M''(-q_t)/M'(-q_t)$, so that we end up with

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -M''\{M^{-1}(S_t)\}I(t),$$

and again the original infinite-dimensional system is reduced to three ODEs.

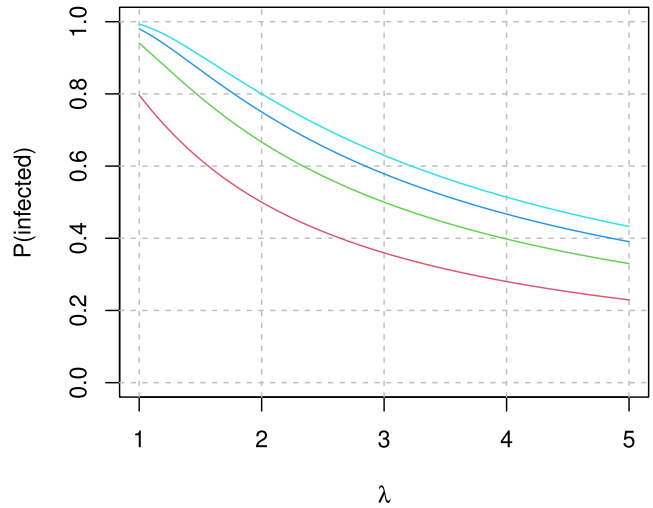



**Figure 8.** Final proportion infected against $\lambda$ for $R_0 = 2, 3, 4, 5$ (ascending, red, green, blue, and turquoise), for Susceptible Exposed Infectious Recovered (SEIR) epidemic models with susceptibility or mixing rates varying across individuals. The age band dependent variability used by Scientific Advisory Group for Emergencies (SAGE) models in the UK corresponds to $\lambda \approx 1.2$. Realistic estimates seem to be in the range 2.5–5.

If $\alpha$ has a gamma$(k, v)$ distribution in the initial susceptible population, with p.d.f. $v^k \alpha^{k-1} e^{-\alpha v}/\Gamma(k)$, then under either model we have

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -R_0 \gamma S_t^{\lambda} I(t)$$

where $R_0$ is the initial pathogen reproductive number and the 'immunity coefficient' $\lambda = 1 + 1/k$, or $1 + 2/k$, for the variable susceptibility or variable contact rate models, respectively. Novozhilov (2008) demonstrates that this is also a good approximation for a wide variety of other initial $\alpha$ distributions. Integration and routine re-arrangement then shows that the final proportion infected, $x$, must satisfy

$$x = 1 - \{1 + (\lambda - 1) R_0 x\}^{-1/(\lambda - 1)}$$

Figure 8 plots the final proportion infected against $\lambda$ for several values of $R_0$. Gomes et al. (2022) estimate that $\lambda = 2.9$ for England and Scotland, and that the heterogeneity with age generally assumed in SAGE modelling corresponds $\lambda \approx 1.2$. Tkachenko et al. (2021) estimates $\lambda$ between 4.1 and 4.7 for several US cities. In the case of transient immunity, subsequent waves are to be expected, of course, but the basic mechanism applies to each of them.

It is possible that this effect was neglected because it was felt to be difficult or impossible to estimate the person-to-person variability in transmission rates, but neglecting an effect known to have a large unidirectional effect on results simply adds spurious precision to estimates that are then almost bound to be wrong. In any case, not including this effect in models calibrated against data is puzzling.

## 4.2 Other major modelling omissions

Another oddity of models used to try to infer $R$, or detect the effect of lockdowns from data, was the fact that they did not model the rather fundamental division of the population into locked-down and key-worker compartments. Those compartments must have different transmission rates, with the difference increasing with the efficacy of lockdown as a suppression measure. The decision was presumably made on grounds of simplicity and the lack of data sufficiently

disaggregated to be informative about rate differences between the compartments. In itself, this simplification may be a reasonable judgement call, but it does imply the need for care to ensure that the model formulation retains the flexibility to deal with the consequences of the simplification.

One such consequence concerns the modelling of $R$ after lockdown. $R$ measures the average number of new infections caused by each existing infection. Crucially, it is the *population of infections* that is being averaged over, not the population of people. Immediately after lockdown $R$ is depressed in the locked-down population, increasingly so with time as the infectious run out of household members to infect. Meanwhile, the key-worker population maintains a higher $R$. Since initially most infections are in the locked-down compartment, the whole population $R$ initially tracks what is happening in that compartment, but over time an ever-increasing proportion of infections is in the key-worker population, so that over time $R$ drifts upwards towards the key-worker compartment $R$. To avoid artefacts, models without separate locked-down and key-worker compartments clearly need to include an $R$ model flexible enough to capture this expected dip and recovery in $R$. Birrell et al. (2021) did this, but the highly cited Flaxman et al. (2020) did not, instead assuming that $R$ was constant during lockdown: the serious artefacts that this induced are discussed in Section 5.3.

Nosocomial infection (disease transmission in hospitals) was also absent from the models despite Wang et al. (2020) reporting a suspected 41% nosocomial infection rate in Wuhan as a key finding in early February 2020, a feature that would be repeated in the first wave in Lombardy in Italy where Boccia et al. (2020) note that 'SARS-CoV-2 became largely a nosocomial infection'. Later analysis showed that within Scotland the proportion of serious Covid that was hospital acquired peaked at around 60% (McKeigue et al., 2021). Model-based analyses, whether statistical or not, are likely to be severely compromised if such a significant transmission route is omitted.

All that said, almost certainly the most important omissions were the negative collateral impacts of the interventions. Of course, there are good reasons why these effects were not included in the epidemic models themselves. But failure to put as much effort into assessing the negative side effects of interventions as was devoted to predicting positive Covid reduction effects is likely to have biased decision-making in a manner unlikely to have achieved anything close to minimum practical societal loss. For example, the July 2020 Government report that attempted some quantification of negative impacts of lockdown (DHSC, 2020) did not attempt any quantification of effects beyond 5 years. For shorter timescales, a much more speculative approach (appendix D4 of the report) was employed than for the disease modelling. This produced results very difficult to reconcile with what actually happened post-2008, as recorded in the data discussed in Section 2.2.

## 5 Lockdowns

That the extreme reduction in contact rates accompanying lockdowns would suppress transmission rates and likely lead infections to decline rather than increase is uncontroversial. However, the retrospective belief that lockdowns were *necessary* for infection levels to fall is a conclusion that seems to be based on informal reasoning and a priori modelling rather than data. The informal reasoning is approximately as follows:

> Across a large number of countries the same pattern was always seen: cases and deaths were increasing until the government imposed a full stay at home lockdown. Only then did cases and deaths decline. Clearly lockdowns caused the decrease, where all preceding measures had failed.

This argument is flawed. Full lockdowns are drastic measures of last resort. As such, no government would impose them unless cases and/or deaths were still increasing, and they were necessarily the last measure imposed. But cases and deaths had to decline eventually. In consequence, the pattern of increase-lockdown-decrease is simply inevitable and conveys no information about lockdowns' role in reversing waves of infection, no matter how often the pattern is repeated. The view that nothing preceding lockdown had worked, because cases and deaths were still increasing until lockdown, neglects the fact that cases and deaths are lagged data, only occurring around one or more weeks after infection. It is what the *current* daily new infection (incidence) rate is doing that indicates the success or otherwise of *current* measures. *Current* cases and deaths cannot do this.

Incidence is difficult to observe directly, but it is possible to retrospectively infer incidence trajectories consistent with the observed daily deaths from Covid. Criteria, ascertainment fraction and distribution of time from infection to event are clearly understood and relatively constant for deaths (none of these things would be true of cases). The simplest approach uses a basic deconvolution model (Wood, 2020, 2021). Suppose that $y_i$ is the number of Covid deaths occurring on day $t_i$, then

$$\mathbb{E}(y_i) = \sum_{d=0}^{D_i} \exp\{f(t_i - d)\}\pi(d)$$

$f(t)$ is the smooth log fatal incidence rate at day $t$ and $\pi(d)$ is the probability of an infection to death time interval of $d$ days. $D_i$ is the maximum lag from infection to death considered. To promote statistical stability, at the start of the epidemic this may be set to somewhere around 20 days, since the first deaths observed will tend to be from shorter duration disease. $D_i$ then grows at a day per day up to some limit (e.g. 80 days). This approach avoids estimating $f$ over a long initial period where $\exp(f)$ is essentially zero. The model can also be multiplied by a second log cyclic smooth term to deal with the slight weekly cycle in deaths seen in some countries. $y_i$ can be assumed to follow a negative binomial or Poisson distribution. The smooth terms in the model can be represented using cubic splines with smoothing parameters estimated by (Laplace approximate) REML. Assuming smoothness on the log scale mitigates against the possibility of smoothing artefacts driven by rapid changes in absolute incidence. Appendix B provides more detail.

The infection-to-death distribution is available from several sources. The meta-analysis of McAloon et al. (2020) combines studies to provide an estimate of the distribution of time from infection to symptom onset. Verity et al. (2020), Linton et al. (2020), and Wu et al. (2020) all provide estimates of the distribution of time from symptom onset to death while properly accounting for right truncation in the data used. Given relatively small sample sizes in these early studies, Wood (2021) integrated the uncertainty in the distributions into the analysis. The results were also compared to those obtained to fitting a model to CHESS data, although it was not possible to obtain data with nosocomial infections filtered out, so that a mixture model approach was necessary. However, later Pritchard et al. (2020) provided results on time from hospitalization to death, and from symptom onset to hospitalization, for a sample of 24,421 fatal cases across multiple countries (the sample is dominated by wealthy countries), at a point in time at which right truncation was a minor issue. Incorporating the results from McAloon et al. (2020), the corresponding infection to death duration model is $\log(d) \sim N(3.151, 0.469^2)$, and given the large underlying sample size, it is the model used here. Figure 9 compares the various onset to death distributions.

## 5.1 Fatal incidence in England

The reconstructed fatal incidence curve for England is shown in the upper panel of Figure 10, based on NHS England hospital deaths data, with the corresponding log of the pathogen reproductive number $R$ shown below. $R$ can be obtained from the incidence curve by assuming a simple SEIR model as described in Wood (2021). The results for the first lockdown are very similar to those obtained by the beginning of May 2020. Later, two more direct statistical reconstructions of incidence became available. The most direct came from the REACT-2 study (H. Ward et al. 2021). Subjects in the study's random sample of English residents who tested positive for SARS-CoV-2 antibodies were asked when their symptoms started. This provides an estimate of the number of newly symptomatic cases each day, from which incidence can be obtained, by applying the same deconvolution method used with the deaths, or simply by lagging the curve by the mean infection to onset duration (5.8 days according to McAloon et al., 2020). The blue curve in the upper panel of Figure 10 shows the result, digitized from H. Ward et al. (2021). Note the somewhat high estimates very early on—presumably representing misattribution of symptoms from other respiratory ailments to Covid, chiefly among subjects whose Covid infection was asymptomatic or very mild. However, even if these 'background infections' were completely suppressed by prelockdown behaviour, there are not enough of them to alone account for the overall

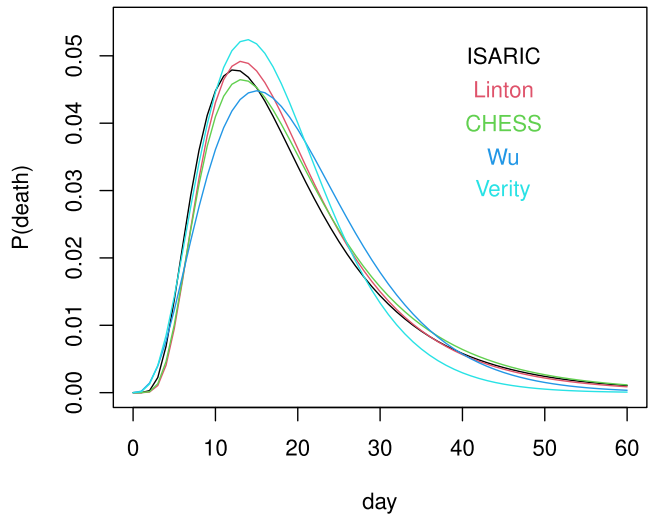



**Figure 9.** Comparison of onset to death duration distributions from various sources. The ISARIC distribution is used here because it is based on the largest sample size by far. Note that the distributions are not significantly different, given the sample sizes involved.

prelockdown drop, without Covid infections also having been in decline. The Office for National Statistics also published incidence reconstructions based on their large-scale randomized surveillance sample from June 2020 (publication was paused for a while in late 2020 and early 2021, while the methods were modified). These are also shown in the upper panel of Figure 10.

Both direct incidence reconstructions align with the death deconvolution approach, albeit both suggesting slightly earlier peaks before each lockdown. There are two possible explanations for these timing mismatches. It could be that the modelled infection to death distributions increase too rapidly at low durations, with the true fatal disease duration distributions being slightly more right-shifted. A less speculative explanation is that deaths occur overwhelmingly in older more vulnerable people, who tend to have lower contact rates and are likely to have reduced these disproportionately relative to the younger healthier population. Subpopulations with lower contact rates peak later than those with higher contact rates, for the reasons discussed in Section 4. So the shift may simply relate to the difference in peak timing expected in a sub-population with lower contact rates.

## 5.2 International comparisons

The correspondence between the direct reconstructions of incidence and the deconvolution of daily deaths strongly suggests that the deconvolution approach is sufficiently reliable to be applied to other countries for which daily death data are available by exact day of death (but surveillance surveys are not). In addition to England, we were able to obtain data for Belgium, Denmark, Italy, the Netherlands, Portugal, Scotland, Spain, Sweden, and Switzerland. The last two are interesting. Both introduced restrictions, but Sweden never introduced full stay-at-home lockdowns, while Switzerland imposed a first lockdown in March 2020, but thereafter remained substantially more open than its neighbours.

The results are shown in Figure 11. Only for the first Belgian and second Italian lockdowns does the turn around in infections coincide with lockdown. For waves at other times and/or locations, the peak in infections precedes lockdown or decline begins without a full lockdown. Although the results imply that the full lockdowns were largely unnecessary for turning around the waves of infection, the reconstructions are consistent with lockdowns having further suppressed infections, causing infection waves to subside more quickly than might otherwise have occurred. In particular, Sweden and Switzerland both experienced broader waves with multiple subsidiary peaks when they did not lock down, a pattern also evident in Italy in the long run-up to its eventual January 2021 lockdown. This suppression of infections is also interesting in light of early model

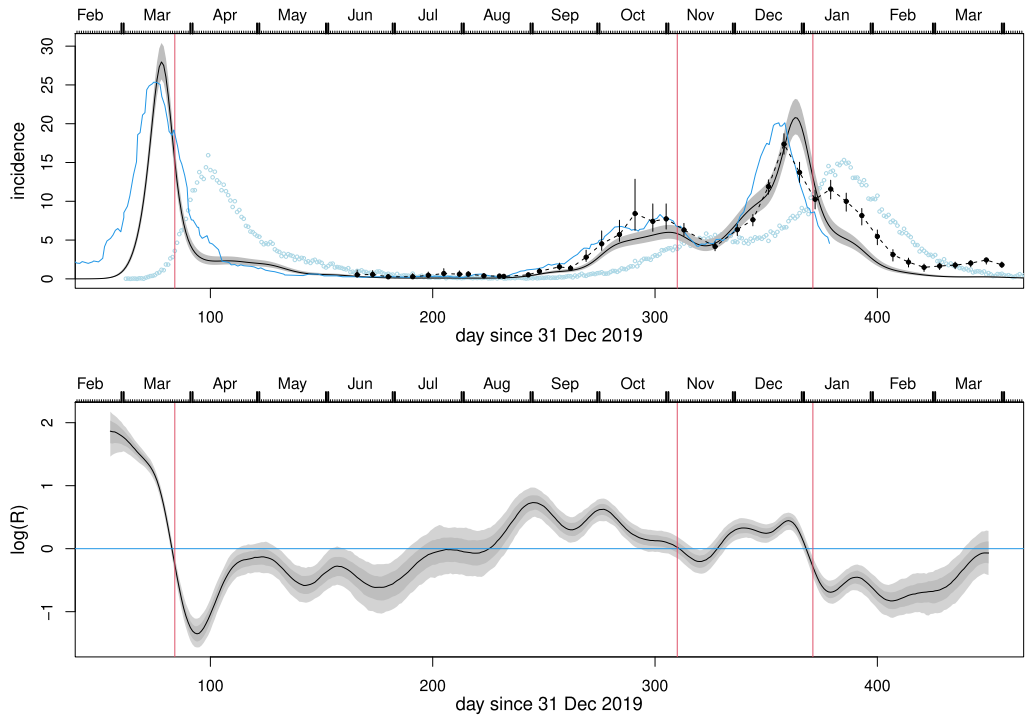



**Figure 10.** Top: The grey bands are 95% credible intervals for fatal incidence (new infections per day) per million, reconstructed from the NHS England daily hospital death data shown as light blue circles. The dark blue curve shows (scaled) reconstructed incidence from the REACT-2 study's random sample of the English population (the study reconstructed incidence of first symptoms, which has been lagged by the 5.8-day average delay from infection to first symptoms). The black dots with confidence bars, joined by dashed lines, are the ONS reconstructions of incidence (scaled) from their randomized surveillance sampling data. Bottom: Natural log of the pathogen reproductive number, $R$, obtained from the grey incidence curve.

predictions that greater suppression of the first wave would delay rather than prevent infections, leading to larger second waves. Such an effect is certainly consistent with the patterns seen for Portugal, Denmark, and Switzerland, and in fact also for Eastern European countries where lockdowns occurred early in the first wave (see, e.g. IFA Mortality Projections Committee, 2023, Chart 4C). From this perspective, the decision to continue suppression well into the summer of 2020 does not seem optimal in terms of health service loading. See online supplementary material for lockdown dates and data sources.

Careful model checking is obviously required when attempting incidence reconstruction. Figure 12 shows the results of repeated 'forward simulations', in which each inferred fatal infection is randomly assigned a duration from the fatal disease duration distribution. This process results in the simulated daily death rates shown as the collection of grey curves on each plot. Overlaid as blue circles are the original raw daily death data, which should look like a plausible draw from the grey curves if the reconstruction is reasonable. Also overlaid are a simple smooth model fit to the daily deaths with CI for the mean. The plots are unproblematic. Note that models were also tried in which $f$ was represented by an adaptive smooth with time varying smoothness, however, these models showed systematic evidence of moderate oversmoothing, presumably related to the rather limited information from which to estimate the several smoothing parameters required. As an illustration of the importance of such model checking, note that at least one group advising UK policymakers attempted to infer incidence by moving each death back in time according to a random draw from the infection to death distribution. The approach is fundamentally flawed as disease duration is not independent of time of death (e.g. at the start of the epidemic, deaths are predominantly from people who had short duration diseases). Forward simulation checks of such a method immediately indicate a problem.

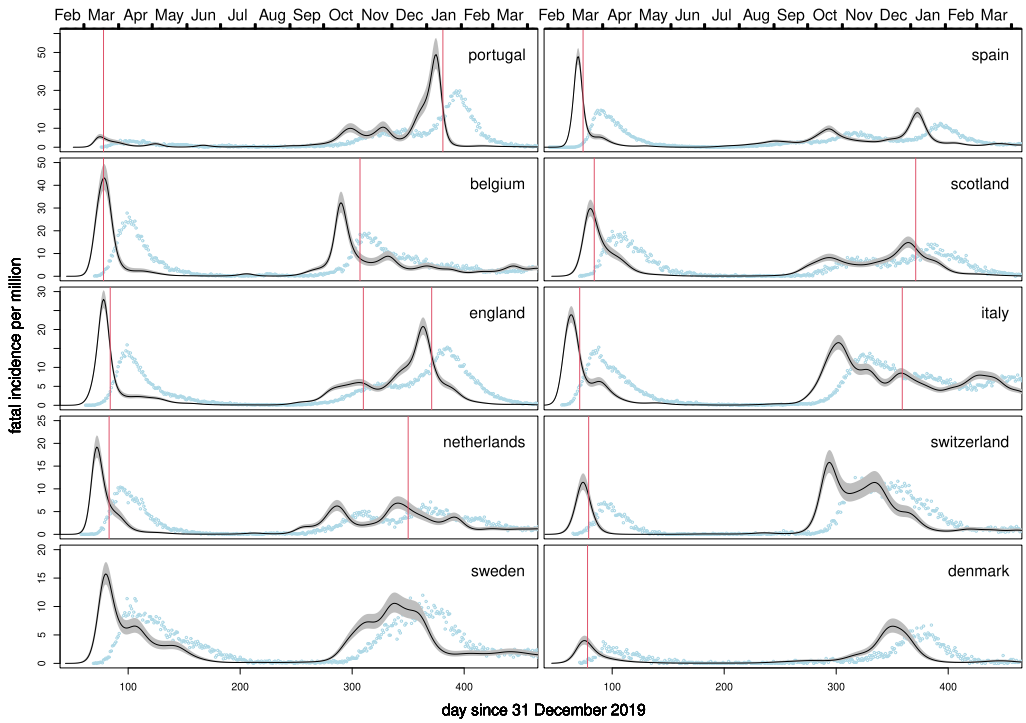



**Figure 11.** Reconstructed fatal incidence for the 10 countries for which reliable daily Covid deaths data by exact day of death were available. Grey confidence bands show reconstructed fatal incidence per million population, with blue circles being the daily deaths per million from which they are obtained. Vertical red lines mark the first day of full national stay-at-home lockdowns.

Wood (2020, 2021) also included detailed checking of the possibility that the smoothness assumptions in the model might cause mistiming in the presence of surging incidence followed by a lockdown-induced collapse—the dominant narrative when the work was undertaken (although now undermined by REACT-2 results). The checking suggested that the timing results were robust.

If one is prepared to accept a simple SEIR model as adequate to describe the aggregate epidemic dynamics in a country, then the incidence reconstructions can be converted to equivalent $R$ reconstructions, as shown in Figure 13. There appear to be no cases for which $R$ had not already declined sharply before lockdown. Only before the first Belgian lockdown was $R$ still appreciably higher than 1, while the second Belgian lockdown apparently came into force when $R$ was already at a low point not seen subsequently. In contrast, at the first lockdown the Netherlands already had $R$ well below 1, but otherwise $R$ was typically around 1 at each country's lockdowns. Only England, Italy, and the Netherlands have $R < 1$ clearly before the first lockdown.

## 5.3 The alternative lockdown narrative

In the UK, analyses from Imperial College (Flaxman et al., 2020; Knock et al., 2020, 2021), and the MRC unit in Cambridge (Birrell et al., 2021) were widely covered and highly influential in promoting the idea that lockdown was the essential component in turning around the first wave of infection. The analyses fitted epidemic models to daily death data, and to other clinical data streams in the case of Knock et al. (2020, 2021). All apparently showed surging incidence up until the eve of the first lockdown, as shown by the red, orange, and pink bands in Figure 14.

The Flaxman et al. (2020) paper attempted to fit a simple renewal model to death data from multiple European countries, assuming that different NPIs had the same multiplicative affect on transmission in each country, irrespective of their order of application, except for the full lockdown effect, which was allowed more country to country variability. To allow for the fact that

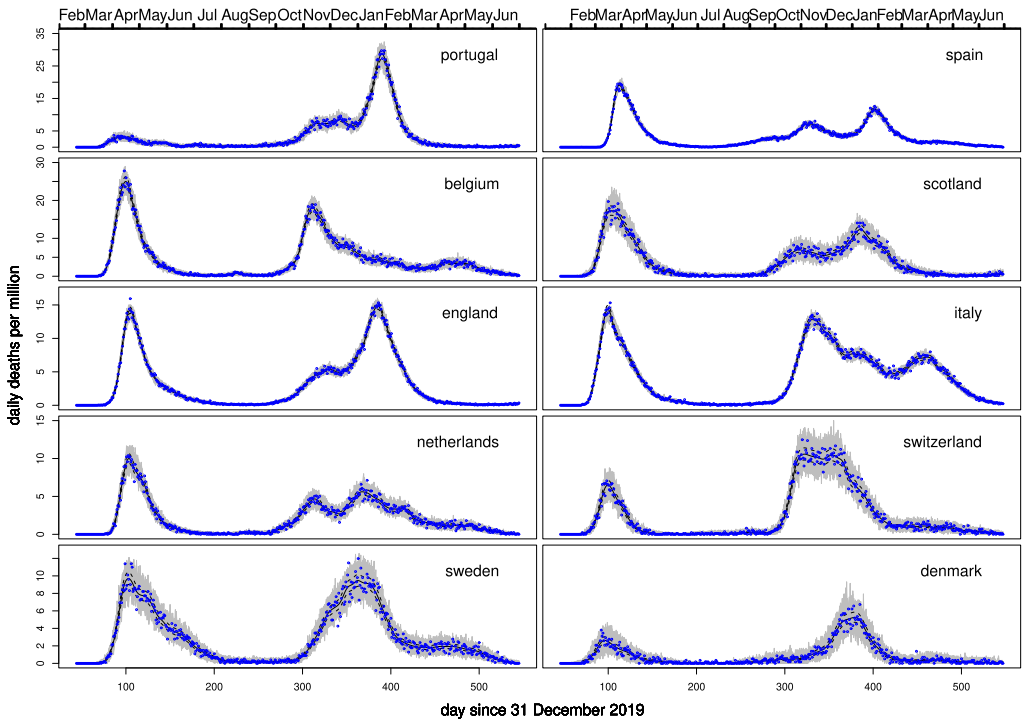



**Figure 12.** Checking plots. Each inferred fatal infection was randomly assigned a duration from the fatal disease duration distribution to produce a daily death curve implied by the fitted model. This simulation process was repeated 100 times to give the grey curves, which can be compared to the observed daily death data plotted in blue.

Sweden did not lock down, the final intervention in Sweden was modelled as if it was lockdown. The approach has been widely criticized (see, e.g. Chin et al., 2021). A particularly insidious problem is the model's treatment of $R$ after full lockdown: $R$ was modelled as a step function, changing only when government policy changed, so constant after lockdown. However, the basic statistical reasoning detailed in Section 4.2 shows that the average $R$ cannot be constant after lockdown, if lockdown is effective at reducing transmission rates. Instead, after an initial decline, a recovery is expected. Exactly this effect is seen in Figure 10, but is precluded by the analysis model of Flaxman et al. (2020). It is difficult to reason intuitively about the consequences of such a structural problem for a highly nonlinear model, so Wood (2021, see also Appendix B) reimplemented the Flaxman et al. (2020) model for England, with the restrictive step function replaced by a cubic spline model for $\log(R)$. The results then match Figure 10, as the blue bands in Figure 14 show.

Birrell et al. (2021) also reported surging incidence up until lockdown (pink in Figure 14), based on an epidemic model fitted to death data. In this case, $R$ was controlled by a contact rate modifier step function with weekly steps, except in the period before lockdown, where it was constant. In other words, increasing incidence until lockdown was simply built into the model.

Knock et al. (2020) fitted an age-structured multi-compartment model to health service death and hospital occupancy data, alongside PCR (polymerase chain reaction) and antibody testing data. The model had some 700 state variables, but inference employed particle filtering with only 96 particles (doubled for the eventually published Knock et al., 2021). For the seven English health service regions, they again purported to show that incidence was increasing and $R > 1$ right up to the eve of the first lockdown (red in Figure 14). In this case, $R$ was controlled by a piecewise linear contact rate modifier with 12 knots at selected government intervention points. Again, there is insufficient flexibility to capture the postlockdown dip and recovery in $R$ expected if lockdown reduces contact rates. Wood & Wit (2021, see also Appendix B) replicated the analysis, replacing the contact rate modifier with an adaptive spline and resetting several rate constants to the values given in the literature cited by Knock et al. (2020) as their source. A simpler

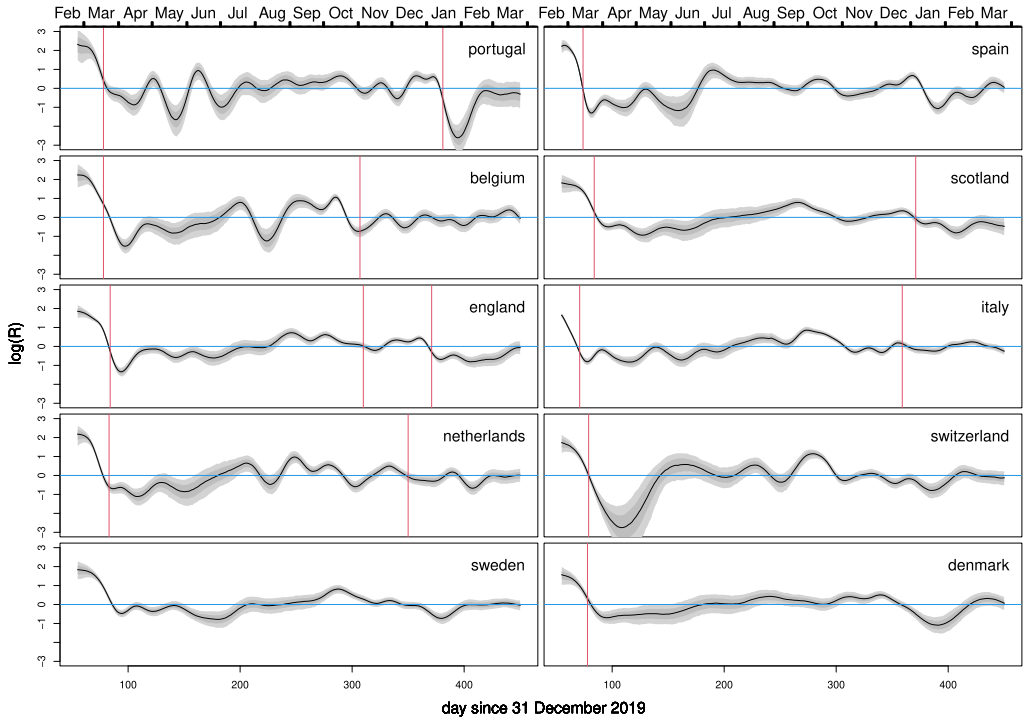



**Figure 13.** The log ($R$) trajectory required for a simple Susceptible Exposed Infectious Recovered (SEIR) model to produce the reconstructed incidence curves, assuming a mean time to infectivity of 3 days and a mean infectious duration of 5 days. Again, vertical red lines mark the lockdown dates. The reconstructions are conditional on an SEIR model structure being reasonable, which it certainly is not at very low incidence rates: hence rapid fluctuations during periods of low incidence are unlikely to be meaningful.

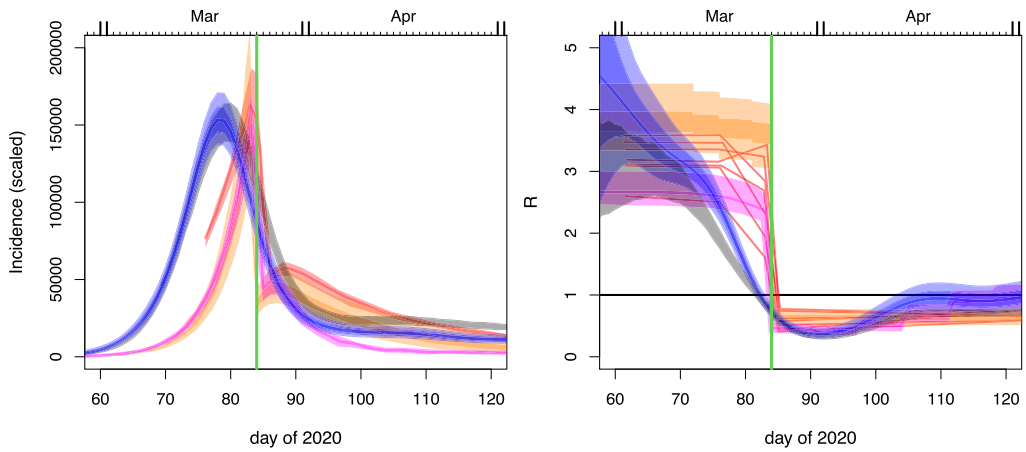



**Figure 14.** Left: incidence reconstructions (vertically scaled for plotting). Right: $R$ reconstructions. The green vertical line is the first day of lockdown. Red: IC Knock et al. (2020) ($R$ separate for each region); Orange: IC Flaxman et al. (2020); Pink: MRC for London Birrell et al. (2021); Grey: replication of Knock et al. (2020) by Wood and Wit (2021) relaxing restrictive assumptions on $R$; Blue: replication of Flaxman et al. (2020) from Wood (2021), relaxing the restrictive assumptions on $R$.

model estimation scheme was used in place of particle filtering. Again, on relaxation of the strong and unrealistic assumptions on contact rate changes, the results aligned with Figure 10, as the grey bands in Figure 14 show.

The other major plank of the narrative of lockdown necessity, was the fact that mathematical models had predicted that lockdown was essential to turn around infection waves, and that after lockdowns were imposed infections indeed declined. The models were not validated for prediction in advance, and the fact that they were able to predict that Covid spread could be massively reduced by suppressing human contact to the maximum extent possible is an especially undemanding check of model sanity.

Two more discerning tests were available. The first one consists of forward mortality predictions. Early in 2020, Imperial College published a study giving the number of predicted deaths likely to accrue under different social distancing scenarios for a number of countries. Walker et al. (2020) predicted about 35,000 first wave Covid deaths for Sweden under the 'social distancing of the whole population' scenario, short of full lockdown, which is the closest scenario to what Sweden actually did. This is interesting as it represented the only first-wave test of the models' ability to predict what might happen without lockdown. Sweden, in fact, experienced fewer than 6,000 first-wave deaths.

The second test of the models' ability to predict what would happen in the absence of lockdowns came with the omicron variant at the end of 2021. The UK government's SAGE relied heavily on the SPI-M committee which synthesised modelling work on Covid to advise on policy. On 19th December 2021, it issued advice on the omicron variant (SPI-M, 2021), strongly suggesting the urgent need for a fourth lockdown. The following edited extract from the summary gives a flavour of the advice.

> … A key consideration for decision-making is how to avert unsustainable pressure on health and care settings… If the coming wave rises comparatively slowly, then a short intervention for, say, a few weeks can prolong the wave's duration and reduce its peak so that admissions and hospital occupancy remain below levels that would compromise quality of care… enacting an intervention early would give time to detect whether such an intervention is insufficient to avoid a compromise of quality of care and adjust accordingly. If measures are implemented only later … measures would need to be in place for longer and might be too late to avert very high admissions…

The detailed advice was based on modelling from the London School of Hygiene and Tropical Medicine (Barnard et al., 2021) and the University of Warwick (Keeling et al., 2021), with the former given substantially more prominence. The government declined to lock down again, so it is possible to compare model projections with reality. Figure 15 does this. In fact the Warwick modelling, as well as presenting a scenario much worse than the one shown, also showed a scenario under the assumption that omicron was only 10% as severe as the delta variant. The lower part of the interval for this scenario does include what actually happened, but it is fair to say that the SPI-M advice did not present this scenario as one that was credible.

## 6 Discussion

The response to Covid was extraordinary in the extent to which it took place online. Initially in the intense pressure in favour of locking down that built on social media, and then in the movement of so much human and scientific interaction online, once lockdowns and other social distancing measures had been implemented. The tendency for online interactions to polarize, amplify exaggeration, and rapidly promote fashions of thought, panics and enthusiasms, while encouraging availability, and confirmation bias (e.g. Kahneman, 2011), has been well-documented by social commentators (e.g. Zuboff, 2019). These tendencies probably make the online world a less than ideal forum for the careful weighing of evidence. They may also serve to promote an adversarial approach to scientific questions, in which the scientist acts as an advocate, whose role is to marshal the data and arguments supporting their theory, rather than as a more neutral interrogator of what data may reveal about reality. The adversarial approach may have advantages, when time is not pressing and there are opposing advocates to attempt falsification, but is perhaps less suited to an emergency, especially if opposing views are characterized as presenting a danger to public health, for example.

The question of lockdown's necessity in turning around waves of infection provides an example where the most careful evidence weighing was appropriate, given the profoundly damaging nature of the intervention. Whatever one's views about how risks should have been balanced in the initial

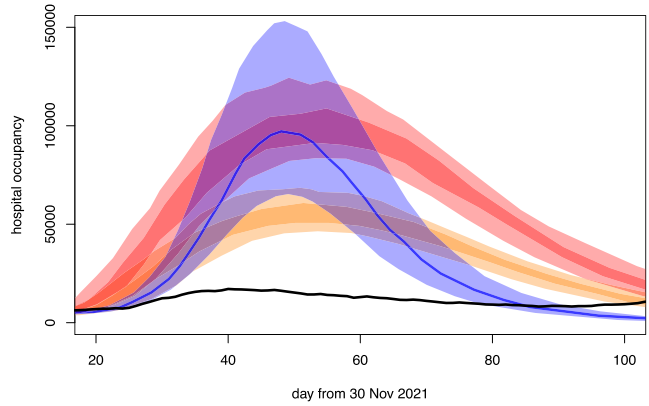



**Figure 15.** LSHTM projections for NHS England Covid hospital bed occupancy as a result of omicron, without lockdown in orange and red. These are the scenarios that are most optimistic about booster efficacy, with the orange representing low vaccine escape assumptions and the red high vaccine escape. In blue are Warwick projections assuming omicron to be 50% as severe as delta. This is the modelling used in the 19th December SPI-M advice. The black line is actual occupancy. Scenarios digitized from the source documents. Actual occupancy figures are NHS England data.

decision to lock down, there was surely an urgent need for rapid and clear-eyed evaluation of whether this experimental intervention had, in fact, been necessary, as soon as possible after its imposition. Instead, models appear to have been treated as evidence and informal intuitive reasoning preferred to what the data strongly implied. In part, the reliance on models may reflect a confusion between updating beliefs and validating them: the notion that updating the distribution of a model's parameters using data in some way validates the model, which it does only in the limited sense of failing to immediately falsify it. The models appearing to indicate the necessity of lockdown can indeed be updated/fitted using data, but highly nonlinear models often have the beguiling flexibility to be able to reproduce a wide range of data, quite irrespective of how their structure reflects the real data generating mechanism. One could argue that the media's response to this issue also tended to confuse the majority opinion of scientists with scientific evidence, emphasising perhaps that expertise is a good reason for listening carefully to an expert's argument, but not for accepting it.

The detachment from external objective reality promoted by the online environment provides fertile ground for the postmodern idea that language, in fact, creates, controls or *is* reality, at least in the social sphere, perhaps feeding the cognate notion that mathematics, the language of science, is equivalent to or controls scientific reality. The 'illusion of control' that this creates (e.g. Gupta, 2001) is seductive, but unhelpful if it results in an excessive effort being devoted to modelling rather than measurement. This is not an argument against quantitative science. For example, the scientifically advised central UK government did not fall for the deeply unscientific and innumerate belief (see Dowdle, 1998) that an endemic disease could, for the first time in history, be eliminated by physical distancing measures, if only these were sufficiently stringent and prolonged ('zero Covid'). Sticking to the quantitative science, in this respect, avoided the even greater collateral damage of harsher, more prolonged measures.

But there was room for improvement in the balance between theory and data, and between modelling and measurement. The public availability of data was in many respects exemplary, with the ONS providing solid evidence on many aspects of the crisis, along with studies such as ISARIC and REACT. But other data was effectively closed to general scrutiny. Data on nosocomial infection was particularly closely guarded, and sensitivity over this appeared to also limit the availability of some other data, such as that relating to the time from first symptoms to death. The ability to independently check and replicate the modelling used to advice policy is severely limited if such data are restricted to an inner circle of advisors. Academic statisticians cannot do their job, of thinking critically about data, unless the data are accessible. Similarly, while some modellers, such as the Imperial College group, took repeatability seriously enough that replication was possible (providing data, comprehensive statement of models and code), other models used by SPI-M were impractical to replicate given what was provided.

Another obvious area of concern is the length of time that it took for randomized surveillance sampling to get underway, given that PCR tests were available from January 2020, when the first UK Covid cases were confirmed (UK government figures put PCR test processing capacity at over 6,000 per day by March 20th, up from 1,500 per day on March 11th. With standard statistical methods for batch testing, a fraction of that capacity would suffice for a useful survey, and the labour force survey sampling frame was already available). In a situation so serious that almost the whole population could be confined to their own homes for 23 hr per day without external in-person contact, it seems incongruous that the first surveillance samples to *measure* the actual state of the epidemic were not taken until 25 April 2020, nearly 3 months after the first UK Covid cases, 2 months after Lombardy and 7 weeks after the first UK death. It seems unlikely that deficiencies in a rapidly rolled out survey, refined as it progressed, could have been worse than not having surveillance data.

A further problem is the limited role that the collateral risks from the measures seem to have played in decision-making. At the very least, a 'red team' with similar heft to SAGE would seem appropriate as soon as a massively costly experimental intervention becomes a serious possibility. Such a team might, for example, have questioned the fairness of allowing the cost per life year saved from Covid to be many times the usual NICE threshold for approval of an intervention (about £30,000 per QALY).

Some aspects of the combination of evidence also appear to have been less than ideal. For example, the 19 December 2021 SPI-M statement on omicron (see Section 5.3) creates the worrying impression that the Warwick modelling may have been somewhat down-weighted in the advice because of the very wide range of outcomes that it suggested were possible, with the LSHTM model given more prominence as a result of its apparently lower uncertainty. But the LSHTM model's increased precision was an illusion created in large part by neglecting the very wide uncertainty in the relationship between vaccine efficacy at blocking transmission and hospitalization used in the modelling.

Also worth discussion is the SPI-M insistence that they were presenting projections (or scenarios), but not predictions, while at the same time making probability statements about them. This seems, at best, philosophically awkward. How is a policy maker to interpret a probability statement about a projection if it is not a prediction? Perhaps 'if the world is sufficiently like the model, then this is the probability of the event of interest'? To act on such a probability, the policy-maker then needs to have some reasonable idea of the probability that the world is like the model. But to declare that projections are not predictions is to declare that this latter probability is unknown. We think that the resolution of this problem probably leads back to the need to statistically validate models for prediction. Otherwise, basing health interventions on model predictions seems worryingly close to licensing a new drug without a clinical trial.

Finally, what was the strategy for dealing with the eventuality that an effective vaccine could not be developed? Without such a strategy, it is difficult to see that risk was, in fact, being managed, however matters eventually turned out. If there was a strategy, but it was not publicly discussed, it is difficult to see it as having the democratic legitimacy one might expect in an open society.

## Acknowledgments

We are especially grateful to Nancy Reid for suggesting that this paper be written. It would not have been otherwise. Thanks also to Jonathan Rougier, Peter Green, Nicole Augustin, Matteo Fasiolo, Helen Colhoun, and Dan Coggan for various discussions of some of the issues raised here, and to the referees for a wealth of useful comments and for providing a number of extra references. S.N.W. is also grateful for help from and discussions with a number of medics who cannot be named, and Dr Alistair Montgomery who can be.

*Conflicts of interest:* None declared.

## Data availability

The data used in the paper are provided in the supplementary material.

## Supplementary material

Supplementary material is available online at *Journal of the Royal Statistical Society: Series A*.

## Appendix A. Iterating life table demography

This appendix provides some details on the iteration of ageing and deaths used in Section 2.4.

Let $y_i$ denote the deaths in week $w_i$ of the year, corresponding to time $t_i$ since the start of the data. To estimate the annual cycle in death rates the generalized additive model

$$\mu_i = f_1(w_i) + f_2(t_i), \qquad \frac{y_i - \mu_i}{\sigma} \sim t_\nu$$

was estimated from 2017 to 2019 data, where $f_1$ is a cyclic smooth function and $f_2$ is a centred slowly varying smooth function, while $\sigma$ and $\nu$ are parameters to be estimated. Then, $d_w = \hat{f}_1(w)/\sum_{w=1}^{52} \hat{f}_1(w)$ defines the multiplier of average weekly mortality required to account for seasonal variation.

The UK population at the start of 2017 or 2020 is available in 1-year age classes, from 0 to 99 plus a '100+' class. All 1-year age class populations were then split into 52 one-week age classes. This was done by fitting a monotonic interpolating spline to the annual cumulative population by age data, and then simply differencing the resulting fit to obtain weekly populations. In this way, the weekly populations vary smoothly, without year-end discontinuity, while the total for each year exactly matches the original yearly data. The approach neglects seasonal birth rate fluctuations. The population in the first week age class is also taken to be the weekly birth rate (the crudeness of this approximation having negligible impact on total deaths). Note that the method also works with data aggregated more coarsely than by yearly age classes.

The life tables provide instantaneous per capita death rates $m_a$ (units year$^{-1}$) for each 1 year age group $a = 0, \ldots, 100$. The average proportion of the age group then dying in 1 week is $q_a = 1 - \exp(-m_a/52)$. Hence, the proportion of 1-year age group $a$ dying in week of year $w$ is $q_a d_w$. Given these preliminaries, the demography is iterated forward using a weekly time step in which the expected deaths are subtracted from the population in each weekly age class before each class is shifted onwards 1 week, and new births are added to the first age class. The per capita mortality rate in a weekly age class is taken as the mortality in its corresponding yearly age class.

When iterated for 3 years from the estimated population by age at the start of 2017, this approach slightly underestimates actual deaths by just under 50 out of 1.8 million (<0.0026%). This slight underestimation will lead to a slight overestimation in excess deaths (somewhere around 0.05%). Code and data used are provided in the online supplementary material.

## Appendix B. Modelling with smooth functions

Both the death deconvolution models and the replications of Flaxman et al. (2020) and Knock et al. (2020), covered in Section 5, are statistical models in which smooth functions of time are to be estimated alongside other parameters. A basis expansion is employed for the smooth function, $f(t) = \sum_{k=1}^{K} \beta_k b_k(t)$ where $\beta_k$ is an unknown coefficient targeted by statistical inference and $b_k(t)$ a known basis function chosen for good approximation theoretic properties. A cubic spline basis is convenient. Associated with $f(t)$ is a smoothing penalty, such as $\lambda \int f''(t)^2 dt = \lambda \boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta}$ ($\mathbf{S}$ known), which can be used to penalize complexity of $f$ during inference, tuneably via the smoothing parameter $\lambda$. In a Bayesian setting, it is natural to view such a penalty as being induced by an improper Gaussian smoothing prior $\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{S}^-/\lambda)$.

Denoting the model log-likelihood as $l$ and expanding $\boldsymbol{\beta}$ to include any other model parameters (and zero padding $\mathbf{S}$ accordingly), then the maximum penalized likelihood estimates of $\boldsymbol{\beta}$ are given by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\, l(\boldsymbol{\beta}) - \frac{\lambda}{2} \boldsymbol{\beta}^T \mathbf{S} \boldsymbol{\beta}, \tag{B1}$$

which is also the posterior mode under the Bayesian view. Pushing the Bayesian view further gives the large sample approximation

$$\boldsymbol{\beta} \mid \mathbf{y} \sim N(\hat{\boldsymbol{\beta}}, \mathbf{V}_\beta)$$

where $\mathbf{V}_\beta = (-\partial^2 l/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T + \lambda\mathbf{S})^{-1}$. Writing $\pi_G(\boldsymbol{\beta} \mid \mathbf{y})$ for this Gaussian approximation to the posterior and $\pi(\boldsymbol{\beta})$ for the smoothing prior, the marginal likelihood is approximately $\exp\{l(\hat{\boldsymbol{\beta}})\}\pi(\hat{\boldsymbol{\beta}})/\pi_G(\hat{\boldsymbol{\beta}} \mid \mathbf{y})$ (Laplace approximation) which can be maximized to estimate $\lambda$. This approach is equally applicable when there are several smoothing parameters and a penalty of the form $\boldsymbol{\beta}^T\mathbf{S}_\lambda\boldsymbol{\beta}$ where $\mathbf{S}_\lambda = \sum \lambda_j \mathbf{S}_j$.

Direct approximate marginal likelihood maximisation would involve nested optimization, which can be tedious to implement for a bespoke dynamic model. However, Wood and Fasiolo (2017) demonstrate how it can be approximately optimized using a simple iteration that alternates Newton optimization of (B1) with updates

$$\lambda_j \leftarrow \frac{\mathrm{tr}(\mathbf{S}_\lambda^- \mathbf{S}_j) - \mathrm{tr}(\mathbf{V}_\beta \mathbf{S}_j)}{\hat{\boldsymbol{\beta}}^T \mathbf{S}_j \hat{\boldsymbol{\beta}}} \lambda_j$$

(note that for many penalties $\mathrm{tr}(\mathbf{S}_\lambda^- \mathbf{S}_j) = \mathrm{rank}(\mathbf{S}_j)/\lambda_j$.) Hence, inference about $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$ then requires only the first and second derivatives of the model log-likelihood with respect to $\boldsymbol{\beta}$. For the deconvolution model, these are straightforward to obtain. For the renewal model of Flaxman et al. (2020), an iterative system for the first and second derivatives of the (discrete-time) dynamic model is required but is relatively straightforward to produce.

For the Knock et al. (2020) ODE model, a system of 'sensitivity' ODEs has to be solved in order to compute first derivatives—this involves 10s of thousands of ODEs, which, while inconvenient, turns out to be numerically inexpensive. However, the second derivative system is impractical. Fortunately, there is an alternative. Solving (B1) by quasi-Newton, the Hessian of the log-likelihood required to compute $\mathbf{V}_\beta$ can be obtained by finite differencing the numerically exact first derivatives, in which case the smoothing parameter updates can proceed using the above update formula, with the approximate posterior available 'for free'.

## References


Acheson D. (1998). *Independent inquiry into inequalities in health report*. The Stationery Office.

Amin-Chowdhury Z., & Ladhani S. N. (2021). Causation or confounding: Why controls are critical for characterizing long COVID. *Nature Medicine*, *27*(7), 1129–1130. https://doi.org/10.1038/s41591-021-01402-w

Appelman B., Charlton B. T., Goulding R. P., Kerkhoff T. J., Breedveld E. A., Noort W., Offringa C., Bloemers F. W., van Weeghel M., Schomakers B. V., Coelho P., Posthuma J. J., Aronica E., Joost Wiersinga W., van Vugt M., & Wüst R. C. I. (2024). Muscle abnormalities worsen after post-exertional malaise in long COVID. *Nature Communications*, *15*(1), 1–15. https://doi.org/10.1038/s41467-023-44432-3

Atchison C. J., Davies B., Cooper E., Lound A., Whitaker M., Hampshire A., Azor A., Donnelly C. A., Chadeau-Hyam M., Cooke G. S., Ward H., & Elliott P. (2023). Long-term health impacts of COVID-19 among 242,712 adults in England. *Nature Communications*, *14*(1), 6588. https://doi.org/10.1038/s41467-023-41879-2

Ayoubkhani D., Khunti K., Nafilyan V., Maddox T., Humberstone B., Diamond I., & Banerjee A. (2021). Post-covid syndrome in individuals admitted to hospital with COVID-19: Retrospective cohort study. *BMJ*, *372*, n693. https://doi.org/10.1136/bmj.n693

Ayoubkhani D., Pawelek P., & Gaughan C. (2021). *Technical article: Updated estimates of the prevalence of post-acute symptoms among people with coronavirus (COVID-19) in the UK: 26 April 2020 to 1 August 2021* (Technical Report). Office for National Statistics.

Ballering A. V., van Zon S. K., Olde Hartman T. C., & Rosmalen J. G. (2022). Persistence of somatic symptoms after COVID-19 in the Netherlands: An observational cohort study. *The Lancet*, *400*(10350), 452–461. https://doi.org/10.1016/S0140-6736(22)01214-4

Barnard R. C., Davies N. G., Pearson C. A. B., Jit M., & Edmunds W. J. (2021). *Projected epidemiological consequences of the Omicron SARS-CoV-2 variant in England, December 2021 to April 2022*. medRxiv. https://doi.org/10.1101/2021.12.15.21267858

Behnood S. A., Shafran R., Bennett S., Zhang A. X., O'Mahoney L. L., Stephenson T. J., Ladhani S., De Stavola B. L., Viner R., & Swann O. V. (2022). Persistent symptoms following SARS-CoV-2 infection amongst

children and young people: A meta-analysis of controlled and uncontrolled studies. *Journal of Infection*, *84*(2), 158–170. https://doi.org/10.1016/j.jinf.2021.11.011
Birrell P., Blake J., Van Leeuwen E., Gent N., & De Angelis D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: The first wave. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200279. https://doi.org/10.1098/rstb.2020.0279
Boccia S., Ricciardi W., & Ioannidis J. P. (2020). What other countries can learn from Italy during the COVID-19 pandemic. *JAMA Internal Medicine*, *180*(7), 927–928. https://doi.org/10.1001/jamainternmed.2020.1447
Brainard J., Jones N. R., Lake I. R., Hooper L., & Hunter P. R. (2020). Community use of face masks and similar barriers to prevent respiratory illness such as COVID-19: A rapid scoping review. *Euro Surveillance*, *25*(49), 2000725. https://doi.org/10.2807/1560-7917.ES.2020.25.49.2000725
Brainard J., Lake I. R., Morbey R. A., Jones N. R., Elliot A. J., & Hunter P. R. (2023). Comparison of surveillance systems for monitoring COVID-19 in England: A retrospective observational study. *The Lancet: Public Health*, *8*(11), e850–e858. https://doi.org/10.1016/S2468-2667(23)00219-0
Brown M., Gerrard J., McKinlay L., Marquess J., Sparrow T., & Andrews R. (2023). Ongoing symptoms and functional impairment 12 weeks after testing positive for SARS-CoV-2 or influenza in Australia: An observational cohort study. *BMJ Public Health*, *1*(1), e000060. https://doi.org/10.1136/bmjph-2023-000060
Case A., & Deaton A. (2017). Mortality and morbidity in the 21st century. *Brookings Papers on Economic Activity*, *2017*(1), 397–476. https://doi.org/10.1353/eca.2017.0005
Chin V., Ioannidis J. P., Tanner M. A., & Cripps S. (2021). Effect estimates of COVID-19 non-pharmaceutical interventions are non-robust and highly model-dependent. *Journal of Clinical Epidemiology*, *136*(7), 96–132. https://doi.org/10.1016/j.jclinepi.2021.03.014
Ciment J. (1999). Life expectancy of Russian men falls to 58. *British Medical Journal*, *319*(7208), 468–468. https://doi.org/10.1136/bmj.319.7208.468a
Cowling B. J., & Leung G. M. (2020). Face masks and COVID-19: Don't let perfect be the enemy of good. *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, *25*(49), 2001998. https://doi.org/10.2807/1560-7917.ES.2020.25.49.2001998
Crafts N. (2023). *The welfare state and-inequality: Were the UK reforms of the 1940s a success?* (Technical Report). IFS: Institute for Fiscal Studies.
DHSC (2020). *Direct and indirect impacts of COVID-19 on excess deaths and morbidity*. Department of Health and Social Care, Office for National Statistics, Government Actuary's Department and Home Office. https://www.gov.uk/government/publications/dhsconsgadho-direct-and-indirect-impacts-of-covid-19-on-excess-deaths-and-morbidity-15-july-2020.
Dowdle W. R. (1998). The principles of disease elimination and eradication. *Bulletin-world Health Organization*, *76*, 22–25. https://iris.who.int/bitstream/handle/10665/260633/PMC2305684.pdf
Ferguson N. M., Laydon D., Nedjati-Gilani G., Imai N., Ainslie K., Baguelin M., Bhatia S., Boonyasiri A., Cucunubá Z., Cuomo-Dannenburg G., Dighe A., Dorigatti I., Fu H., Gaythorpe K., Green W., Hamlet A., Hinsley W., Okell L. C., van Elsland S., … Ghani A. C. (2020). *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand* (Technical Report). Imperial College COVID-19 Response Team.
Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., Monod M., Imperial College COVID-19 Response Team, Ghani A. C., Donnelly C. A., Riley S., Vollmer M. A. C., Ferguson N. M., Okell L. C., & Bhatt S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7
Gomes M. G. M., Ferreira M. U., Corder R. M., King J. G., Souto-Maior C., Penha-Gonçalves C., Gonçalves G., Chikina M., Pegden W., & Aguas R. (2022). Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *Journal of Theoretical Biology*, *540*, 111063. https://doi.org/10.1016/j.jtbi.2022.111063
Gray A. M. (1982). Inequalities in health. The black report: A summary and comment. *International Journal of Health Services*, *12*(3), 349–380. https://doi.org/10.2190/XXMM-JMQU-2A7Y-HX1E
Gupta S. (2001). Avoiding ambiguity. *Nature*, *412*(6847), 589. https://doi.org/10.1038/35088152
Hall R. E. (2009). *Inflation: Causes and effects*. University of Chicago Press.
Hanlon P., Chadwick F., Shah A., Wood R., Minton J., McCartney G., Fischbacher C., Mair F. S., Husmeier D., & Matthiopoulos J. (2020). *COVID-19–exploring the implications of long-term condition type and extent of multimorbidity on years of life lost: A modelling study*. https://wellcomeopenresearch.org/articles/5-75.
Haslam A., & Prasad V. (2023). Comparability of control and comparison groups in studies assessing long COVID. *The American Journal of Medicine*, *138*(1), 148–152. https://doi.org/10.1016/j.amjmed.2023.01.005
Hastie C. E., Lowe D. J., McAuley A., Winter A. J., Mills N. L., Black C., Scott J. T., O'Donnell C. A., Blane D. N., Browne S., Ibbotson T. R., & Pell J. P. (2022). Outcomes among confirmed cases and a matched comparison group in the Long-COVID in Scotland study. *Nature Communications*, *13*(1), 5663. https://doi.org/10.1038/s41467-022-33415-5

Herridge M. S. (2011). Recovery and long-term outcome in acute respiratory distress syndrome. *Critical Care Clinics*, *27*(3), 685–704. https://doi.org/10.1016/j.ccc.2011.04.003

Høeg T. B., Ladhani S., & Prasad V. (2023). How methodological pitfalls have created widespread misunderstanding about long COVID. *BMJ Evidence-Based Medicine*, *29*, 142–146. https://doi.org/10.1136/bmjebm-2023-112338

Holi M., Sammallahti P., & Aalberg V. (1998). A Finnish validation study of the SCL-90. *Acta Psychiatrica Scandinavica*, *97*(1), 42–46. https://doi.org/10.1111/acp.1998.97.issue-1

Holleyman R. J., Barnard S., Bauer-Staeb C., Hughes A., Dunn S., Fox S., Newton J. N., Fitzpatrick J., Waller Z., Deehan D. J., Charlett A., Gregson C. L., Wilson R., Fryers P., Goldblatt P., & Burton P. (2023). Adjusting expected deaths for mortality displacement during the COVID-19 pandemic: A model based counterfactual approach at the level of individuals. *BMC Medical Research Methodology*, *23*(1), 241. https://doi.org/10.1186/s12874-023-01984-8

Home Office (2021). *Migrant journey: 2020 report* (Technical Report). UK Government.

Hopkins R. O., Weaver L. K., Pope D., Orme Jr J. F., Bigler E. D., & Larson-Lohr V. (1999). Neuropsychological sequelae and impaired health status in survivors of severe acute respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine*, *160*(1), 50–56. https://doi.org/10.1164/ajrccm.160.1.9708059

Huang C., Wang Y., Li X., Ren L., Zhao J., Hu Y., Zhang L., Fan G., Xu J., Gu X., Cheng Z., Yu T., Xia J., Wei Y., Wu W., Xie X., Yin W., Li H., Liu M., … Cao B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Hutton G., & Powell A. (2021). UK supply chain problems. House of Commons Library Research Briefing CBP 9350.

IFA Mortality Projections Committee (2023). *WORKING PAPER 180: International excess mortality* (Technical Report). Institute and Faculty of Actuaries.

Islam N., Jdanov D. A., Shkolnikov V. M., Khunti K., Kawachi I., White M., Lewington S., & Lacey B. (2021). Effects of covid-19 pandemic on life expectancy and premature mortality in 2020: Time series analysis in 37 countries. *BMJ*, *375*. https://doi.org/10.1136/bmj-2021-066768

Kahneman D. (2011). *Thinking, fast and slow*. Penguin.

Keeling M. J., Brooks-Pollock E., Challen R., Danon L., Dyson L., Gog J. R., Rincón L. G., Hill E. M., Pellis L., & Read J. M. (2021). *Short-term projections based on early omicron variant dynamics in England*. medRxiv. https://doi.org/10.1101/2021.12.30.21268307

Knock E. S., Whittles L. K., Lees J. A., Perez Guzman P. N., Verity R., Fitzjohn R. G., Gaythorpe K. A. M., Imai N., Hinsley W., & Okell L. C. (2020). *Report 41: The 2020 SARS-CoV-2 epidemic in England: Key epidemiological drivers and impact of interventions*. Imperial College London. https://doi.org/10.25561/85146

Knock E. S., Whittles L. K., Lees J. A., Perez Guzman P. N., Verity R., Fitzjohn R. G., Gaythorpe K. A. M., Imai N., Hinsley W., Okell L. C., Rosello A., Kantas N., Walters C. E., Bhatia S., Watson O. J., Whittaker C., Cattarino L., Boonyasiri A., Djaafara B. A., … Baguelin M. (2021). Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Science Translational Medicine*, *13*(602), eabg4262. https://doi.org/10.1126/scitranslmed.abg4262

Linton N. M., Kobayashi T., Yang Y., Hayashi K., Akhmetzhanov A. R., Jung S.-m., Yuan B., Kinoshita R., & Nishiura H. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *Journal of Clinical Medicine*, *9*(2), 538. https://doi.org/10.3390/jcm9020538

Liu Y., Morgenstern C., Kelly J., Lowe R., & Jit M. (2021). The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Medicine*, *19*(1), 1–12. https://doi.org/10.1186/s12916-020-01826-0

Major L. E., Eyles A., & Machin S. (2021). *Learning loss since lockdown: Variation across the home nations*. COVID-19 Analysis Series. No. 023. Centre for Economic Performance.

Marmot M., Allen J., Boyce T., Goldblatt P., & Morrison J. (2020). *Health equity in England: The marmot review 10 years on*. The Health Foundation.

Matta J., Wiernik E., Robineau O., Carrat F., Touvier M., Severi G., de Lamballerie X., Blanché H., Deleuze J.-F., Gouraud C., Hoertel N., Ranque B., Goldberg M., Zins M., Lemogne C., & Santé, Pratiques, Relations et Inégalités Sociales en Population Générale Pendant la Crise COVID-19Sérologie (SAPRIS-SERO) Study Group (2022). Association of self-reported COVID-19 infection and SARS-CoV-2 serology test results with persistent physical symptoms among French adults during the COVID-19 pandemic. *JAMA Internal Medicine*, *182*(1), 19–25. https://doi.org/10.1001/jamainternmed.2021.6454

McAloon C., Collins Á., Hunt K., Barber A., Byrne A. W., Butler F., Casey M., Griffin J., Lane E., McEvoy D., Wall P., Green M., O'Grady L., & More S. J. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open*, *10*(8), e039652. https://doi.org/10.1136/bmjopen-2020-039652

McKeigue P. M., McAllister D., Caldwell D., Gribben C., Bishop J., McGurnaghan S. J., Armstrong M., Delvaux J., Colville S., Hutchinson S., Robertson C., Lone N., McMenamin J., Goldberg D., & Colhoun H. M. (2021). Relation of severe COVID-19 in Scotland to transmission-related factors and risk conditions eligible for shielding support: REACT-SCOT case–control study. *BMC Medicine*, *19*(1), 1–13. https://doi.org/10.1186/s12916-020-01826-0

Mitze T., Kosfeld R., Rode J., & Wälde K. (2020). Face masks considerably reduce COVID-19 cases in Germany. *Proceedings of the National Academy of Sciences*, *117*(51), 32293–32301. https://doi.org/10.1073/pnas.2015954117

NICE (2022). *COVID-19 rapid guideline: Managing the long-term effects of COVID-19* (Technical Report). National Institute for Health and Care Excellence (NICE), Scottish Intercollegiate Guidelines Network (SIGN) and Royal College of General Practitioners (RCGP).

Novozhilov A. S. (2008). On the spread of epidemics in a closed heterogeneous population. *Mathematical Biosciences*, *215*(2), 177–185. https://doi.org/10.1016/j.mbs.2008.07.010

O'Connor R. C., Wetherall K., Cleare S., McClelland H., Melson A. J., Niedzwiedz C. L., O'Carroll R. E., O'Connor D. B., Platt S., Scowcroft E., Watson B., Zortea T., Ferguson E., & Robb K. A. (2021). Mental health and well-being during the COVID-19 pandemic: Longitudinal analyses of adults in the UK COVID-19 mental health & wellbeing study. *The British Journal of Psychiatry*, *218*(6), 326–333. https://doi.org/10.1192/bjp.2020.212

O'Mahoney L. L., Routen A., Gillies C., Ekezie W., Welford A., Zhang A., Karamchandani U., Simms-Williams N., Cassambai S., Ardavani A., Wilkinson T. J., Hawthorne G., Curtis F., Kingsnorth A. P., Almaqhawi A., Ward T., Ayoubkhani D., Banerjee A., Calvert M., … Khunti K. (2023). The prevalence and long-term health effects of long COVID among hospitalised and non-hospitalised populations: A systematic review and meta-analysis. *EClinicalMedicine*, *55*(9), 101762. https://doi.org/10.1016/j.eclinm.2022.101762

ONS. (2022). *Self-reported long COVID and labour market outcomes, UK: 2022*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/selfreportedlongcovidandlabourmarketoutcomesuk2022/selfreportedlongcovidandlabourmarketoutcomesuk2022

ONS. (2023). *Rising ill-health and economic inactivity because of long-term sickness, UK: 2019 to 2023*. https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/economicinactivity/articles/risingillhealthandeconomicinactivitybecauseoflongtermsicknessuk/2019to2023

Pamuk E. R. (1985). Social class inequality in mortality from 1921 to 1972 in England and Wales. *Population Studies*, *39*(1), 17–31. https://doi.org/10.1080/0032472031000141256

Pritchard M., Dankwa E. A., Hall M., Baillie J. K., Carson G., Docherty A., Donnelly C. A., Dunning J., Fraser C., & Hardwick H. (2020). *ISARIC clinical data report 4 October 2020*. medRxiv. https://doi.org/10.1101/2020.07.17.20155218

Rea M., Pawelek P., & Ayoubkhani D. (2023). *Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK: 30 March 2023* (Technical Report). Office for National Statistics.

Richardson E., McCartney G., Taulbut M., Douglas M., & Craig N. (2023). Population mortality impacts of the rising cost of living in Scotland: Scenario modelling study. *BMJ Public Health*, *1*(1), e000097. https://doi.org/10.1136/bmjph-2023-000097

Riera R., Bagattini Â. M., Pacheco R. L., Pachito D. V., Roitberg F., & Ilbawi A. (2021). Delays and disruptions in cancer health care due to COVID-19 pandemic: Systematic review. *JCO Global Oncology*, *7*(1), 311–323. https://doi.org/10.1200/GO.20.00639

Rougier J., Sparks R. S. J., Cashman K. V., & Brown S. K. (2018). The global magnitude–frequency relationship for large explosive volcanic eruptions. *Earth and Planetary Science Letters*, *482*(5), 621–629. https://doi.org/10.1016/j.epsl.2017.11.015

Ruminska-Zimny E. (1997). Human poverty in transition economies: Regional overview for HDR 1997. Human Development Report Office, United Nations Development Programme New York.

Selvakumar J., Havdal L. B., Drevvatne M., Brodwall E. M., Lund Berven L., Stiansen-Sonerud T., Einvik G., Leegaard T. M., Tjade T., Michelsen A. E., Mollnes T. E., Lund-Johansen F., Holmøy T., Zetterberg H., Blennow K., Sandler C. X., Cvejic E., Lloyd A. R., & Wyller V. B. B. (2023). Prevalence and characteristics associated with Post-COVID-19 condition among nonhospitalized adolescents and young adults. *JAMA Network Open*, *6*(3), e235763–e235763. https://doi.org/10.1001/jamanetworkopen.2023.5763

Shine C. (2022). *Exploring latest trends and insights from ONS data on the Cost of Living*. https://www.gov.wales/sites/default/files/publications/2022-05/third-sector-statistics-user-panel-28-april-2022-exploring-latest-trends-and-insights-from-ons-data-on-the-cost-of-living.pdf.

SPI-M. (2021). *SPI-M-O Chairs: Statement on COVID-19* (Technical Report). https://www.gov.uk/government/publications/spi-m-o-chairs-statement-on-covid-19-19-december-2021/spi-m-o-chairs-statement-on-covid-19-19-december-2021

Stuckler D., Basu S., & McKee M. (2010). Budget crises, health, and social welfare programmes. *BMJ*, *340*, c3311. https://doi.org/10.1136/bmj.c3311

Tkachenko A. V., Maslov S., Elbanna A., Wong G. N., Weiner Z. J., & Goldenfeld N. (2021). Time-dependent heterogeneity leads to transient suppression of the COVID-19 epidemic, not herd immunity. *Proceedings of the National Academy of Sciences*, *118*(17), e2015972118. https://doi.org/10.1073/pnas.2015972118

Van Driel T., Hilderink P., Hanssen D., De Boer P., Rosmalen J., & Oude Voshaar R. (2018). Assessment of somatization and medically unexplained symptoms in later life. *Assessment*, *25*(3), 374–393. https://doi.org/10.1177/1073191117721740

Verity R., Okell L. C., Dorigatti I., Winskill P., Whittaker C., Imai N., Cuomo-Dannenburg G., Thompson H., Walker P. G. T., Fu H., Dighe A., Griffin J. T., Baguelin M., Bhatia S., Boonyasiri A., Cori A., Cucunubá Z., FitzJohn R., Gaythorpe K., … Ferguson N. M. (2020). Estimates of the severity of coronavirus disease 2019: A model-based analysis. *The Lancet: Infectious Diseases*, *20*(6), 669–677. https://doi.org/10.1016/S1473-3099(20)30243-7

Viboud C., Grais R. F., Lafont B. A., Miller M. A., & Simonsen L. (2005). Multinational impact of the 1968 Hong Kong influenza pandemic: Evidence for a smoldering pandemic. *The Journal of Infectious Diseases*, *192*(2), 233–248. https://doi.org/10.1086/jid.2005.192.issue-2

Walker P., Whittaker C., Watson O., Baguelin M., Ainslie K., Bhatia S., Bhatt S., Boonyasiri A., Boyd O., & Cattarino L. (2020). *Report 12: The global impact of COVID-19 and strategies for mitigation and suppression*. London: Imperial College. https://doi.org/10.25561/77735

Wang D., Hu B., Hu C., Zhu F., Liu X., Zhang J., Wang B., Xiang H., Cheng Z., Xiong Y., Zhao Y., Li Y., Wang X., & Peng Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA*, *323*(11), 1061–1069. https://doi.org/10.1001/jama.2020.1585

Ward H., Cooke G., Whitaker M., Redd R., Eales O., Brown J. C., Collet K., Cooper E., Daunt A., & Jones K. (2021). *REACT-2 Round 5: Increasing prevalence of SARS-CoV-2 antibodies demonstrate impact of the second wave and of vaccine roll-out in England*. medRxiv. https://doi.org/10.1101/2021.02.26.21252512

Wolf M. (2023). Lessons from the great reflation. *Financial Times*.

Wood S. N. (2020). 'Did COVID-19 infections decline before UK lockdown?', arXiv, arXiv:2005.02090. https://doi.org/10.48550/arXiv.2005.02090

Wood S. N. (2021). Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? *Biometrics*, *78*(3), 1127–1140. https://doi.org/10.1111/biom.v78.3

Wood S. N., & Fasiolo M. (2017). A generalized Fellner–Schall method for smoothing parameter optimization with application to Tweedie location, scale and shape models. *Biometrics*, *73*(4), 1071–1081. https://doi.org/10.1111/biom.12666

Wood S. N., & Wit E. C. (2021). Was R<1 before the English lockdowns? On modelling mechanistic detail, causality and inference about Covid-19. *PloS One*, *16*(9), e0257455. https://doi.org/10.1371/journal.pone.0257455

Wood S. N., Wit E. C., Fasiolo M., & Green P. J. (2020). COVID-19 and the difficulty of inferring epidemiological parameters from clinical data. *The Lancet: Infectious Diseases*. https://doi.org/10.1016/S1437-3099(20)30437-0

Wu J. T., Leung K., Bushman M., Kishore N., Niehus R., de Salazar P. M., Cowling B. J., Lipsitch M., & Leung G. M. (2020). Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine*, *26*(4), 506–510. https://doi.org/10.1038/s41591-020-0822-7

Zhou F., Yu T., Du R., Fan G., Liu Y., Liu Z., Xiang J., Wang Y., Song B., Gu X., Guan L., Wei Y., Li H., Wu X., Xu J., Tu S., Zhang Y., Chen H., & Cao B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *The Lancet*, *395*(10229), 1054–1062. https://doi.org/10.1016/S0140-6736(20)30566-3

Zuboff S. (2019). *The age of surveillance capitalism*. Profile Books.

*Journal of the Royal Statistical Society Series A: Statistics in Society*, 2026, **189**, 32–138
**Discussion Paper Contribution**

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
**Series A**
Statistics in Society

# Peter Goldblatt's invited contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Peter Goldblatt**

UCL Institute of Health Equity, Department of Epidemiology & Public Health, UCL, London, UK

*Address for correspondence*: Peter Goldblatt, UCL Institute of Health Equity, Department of Epidemiology & Public Health, UCL, 1-19 Torrington Place, London WC1E 7HB, UK. Email: p.goldblatt@ucl.ac.uk

The article presents a new analysis of the number of excess deaths in England during the COVID-19 pandemic, by taking more account than others of the short-term impact on the expected number of deaths of the 1946 boom in births and by applying a life table methodology to calculate these expected numbers. These have quite separate effects on the calculations, and it is a shame that the authors did not provide a clear analysis of the attributable fractions associated with each change—while the first is a temporary phenomenon, the other would have long-term implications for monitoring.

The authors make a number of claims about other aspects of the handling of the pandemic and its consequences. The evidence base for some of these can be questioned.

1. The authors state that 'The 2020 [Marmot] report includes a forensic, databased investigation of how health outcomes and life expectancy for the more disadvantaged were worsened by the exacerbation of economic deprivation following the financial crisis of 2008 and subsequent government response to it'. In fact, the report makes no mention of the financial crisis and simply analyses the health-related effects of regressive policies brought in by the coalition government, resulting in the rich getting richer at the expense of the 30% least wealthy (Figure 3.27 of the report). Subsequent analyses, based on National Audit Office data, show that reductions in central government funding to local authorities were targeted at areas with pre-existing worst health (Goldblatt et al., 2024). None of these policies can be described as 'exacerbated' by the financial crisis.
2. When the authors argue that 'only some 17 thousand people had only Covid and nothing else recorded on their death certificate. When Covid is only one factor among several in a death, it is statistically naive to expect it to contribute a whole extra death in the excess figures', they misrepresent the conceptualization of what constitutes an excess death as well as international guidance on completion of the medical certificate of death. On the first of these, everyone must die at some point. Thus, conceptually, an excess death is one that occurs earlier than it would otherwise have done in the absence of the specified risk (Cox, 1972). On the second certification, it is recognized that people with severe complications of COVID-19 were more likely to have specific pre-existing health problems (Hippisley-Cox et al., 2021; Nafilyan et al., 2021; Bauer-Staeb et al., 2024). The responsibility of the certifier is to record the sequence leading to death (the complications) and other contributory factors (the pre-existing conditions) (WHO 2022). None of these factors alter the fact that COVID-19 resulted in the death occurring prematurely in most

**Received:** May 16, 2025. **Accepted:** May 19, 2025

cases (Hughes et al., 2025) and therefore being 100% responsible for the whole of what we count as an excess death.

3. The authors argue that 'harvesting.... is where an epidemic pathogen brings forward the deaths of some very frail people by only a few weeks or months'. This is to oversimplify a probabilistic process of displacement whereby the likely timing of death is brought forward in proportion to the level of the individual's pre-existing co-morbidities—while the likely dates of death of those on end of life care are not necessarily brought forward at all, those with chronic but life-shortening conditions are likely to die only years earlier than they would otherwise have done (but more years earlier than someone in perfect health) (Hughes et al., 2025). Thus, the pandemic acts as a classic ill-health selection effect (Fox et al., 1982) and such effects typically take three to five years to largely wear-off. The authors' statement that 'over a period of three years many such people will not appear as excess deaths at all, since their death has only been moved within the time period considered' is doubly misleading. First, as argued above, any premature death is an excess death. While the Office for National Statistics (ONS) and the Office for Health Improvement and Disparities use a period definition to present excess deaths (Comparing numbers of deaths within a time period to those expected in the same period), they only do this for relatively short periods (a week or a month) to avoid failing to count displaced by months or more. Doing this for long periods, without excluding those who have died from the calculation of risk, clearly leads to absurd results. For example, if we consider the extreme example of a 100-year period, then everyone would have died within the period and those dying prematurely would be indistinguishable from those who had a normal lifespan, i.e. there would be no count of excess deaths. Second, as Hughes et al. (2025) show, the median displacement of all deaths in England under the age of 80 exceeded three years, after taking account of co-morbidity and other factors.
4. The authors' argument that lockdowns were unnecessary because 'cases and deaths are lagged data, only occurring around one or more weeks after infection' misses a number of key points in relation to the 'wild stage' of the pandemic in March 2020, in particular. First, cases, complications, and deaths had totally different demographic distributions—infections were more common at working ages, while the risk of complications and deaths following infection largely increased with age. So, preventing the further spread of infection from younger to older people was essential as the National Health Service came under unsustainable pressure, even if the infection peak had passed. The evidence that this was not misguided comes from Ayoubkhani et al. (2021b). They showed that the peak of deaths in care homes came weeks after lockdown. This is thought to be a result of working-age agency care staff spreading infection between care homes during lockdown. Had there been no lockdown, working-age individuals would have continued to infect older individuals in the community and thus extended the decline in community infections (albeit after the peak in deaths), and hence both placed more pressure on the NHS and increased numbers of deaths. What the authors do not consider is the scientific argument for introducing lockdown earlier and the modelled consequences of so doing, based on experience in China.
5. The authors' analysis of long COVID is at best confused. There is a long history of collecting self-reported health in surveys (de Bruin et al., 1996). While this does not accord with clinical data, it can be a strong predictor of mortality (Torres-Collado et al., 2022) and inevitably can be over-reported. However, during the pandemic, long COVID was a poorly understood set of symptoms with variable duration and, therefore, limited clinical tools to replace self-reporting. Now, we are in a better position to understand the aetiology of persistent and severe symptoms that affect those with high levels of exposure to the virus (e.g. NHS frontline staff) and those who needed to be admitted to hospital (Ayoubkhani et al., 2021a; Davis et al., 2023; Walker et al., 2023; Appelman et al., 2024). The authors should not negate the role of self-reporting in the process of gaining a better epidemiological picture of long COVID.
6. The authors' assertion that 'it is virtually impossible to determine [the actual number infected}, except perhaps in the unlikely scenario of real-time, continuous, population-wide testing' is to deny the possibility of obtaining modelled estimates from the rich data available for at least part of the pandemic—for example the ONS/Oxford survey and the period

for which free community-wide testing was implemented. The same issues apply to all diseases that are either asymptomatic or under-diagnosed. They also apply to determining the number of unemployed people. In all cases, modelling techniques are needed to triangulate available sources. For many purposes, it is also sufficient to obtain statistics based on the follow-up of those with a positive test during the period of community testing and build that into a model based on the distribution of the latent period between exposure and a positive test.

*Conflicts of interest:* none declared.

## References


Appelman, B., Charlton, B. T., Goulding, R. P., Kerkhoff, T. J., Breedveld, E. A., Noort, W., Offringa, C., Bloemers, F. W., van Weeghel, M., & Schomakers, B. V. (2024). Muscle abnormalities worsen after post-exertional malaise in long COVID. *Nature Communications*, *15*(1), 17. https://doi.org/10.1038/s41467-023-44432-3

Ayoubkhani, D., Khunti, K., Nafilyan, V., Maddox, T., Humberstone, B., Diamond, I., & Banerjee, A. (2021a). Post-COVID syndrome in individuals admitted to hospital with COVID-19: Retrospective cohort study. *BMJ*, *372*, n693. https://doi.org/10.1136/bmj.n693

Ayoubkhani, D., Nafilyan, V., White, C., Goldblatt, P., Gaughan, C., Blackwell, L., Rogers, N., Banerjee, A., Khunti, K., Glickman, M., Humberstone, B., & Diamond, I. (2021b). Ethnic minority groups in England and Wales—Factors associated with the size and timing of elevated COVID-19 mortality: A retrospective cohort study linking census and death records. *International Journal of Epidemiology*, *49*(6), 1951–1962. https://doi.org/10.1093/ije/dyaa208

Bauer-Staeb, C., Holleyman, R. J., Barnard, S., Hughes, A., Dunn, S., Fox, S., Fitzpatrick, J., Newton, J., Fryers, P., Burton, P., & Goldblatt, P. (2024). Risk of death in England following a positive SARS-CoV-2 test: A retrospective national cohort study (March 2020 to September 2022). *PLoS One*, *19*(10), e0304110. https://doi.org/10.1371/journal.pone.0304110

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *34*(2), 187–202. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x

Davis, H. E., McCorkell, L., Vogel, J. M., & Topol, E. J. (2023). Long COVID: Major findings, mechanisms and recommendations. *Nature Reviews: Microbiology*, *21*(3), 133–146. https://doi.org/10.1038/s41579-022-00846-2

de Bruin, A., Picavet, H. S., & Nossikov, A. (1996). Health interview surveys: Towards international harmonization of methods and instruments. *World Health Organization Regional Regional Publications, European Series*, *58*, i–xiii. 1-161.

Fox, A. J., Goldblatt, P. O., & Adelstein, A. M. (1982). Selection and mortality differentials. *Journal of Epidemiology and Community Health*, *36*(2), 69–79. https://doi.org/10.1136/jech.36.2.69

Goldblatt, P., Callaghan, O., Allen, J., & Porritt, F. (2024) *England's Widening Health Gap: Local Places Falling Behind*. https://www.instituteofhealthequity.org/resources-reports/englands-widening-health-gap-local-places-falling-behind. Date accessed May 16, 2025.

Hippisley-Cox, J., Coupland, C. A., Mehta, N., Keogh, R. H., Diaz-Ordaz, K., Khunti, K., Lyons, R. A., Kee, F., Sheikh, A., Rahman, S., Valabhji, J., Harrison, E. M., Sellen, P., Haq, N., Semple, M. G., Johnson, P. W. M., Hayward, A., & Nguyen-Van-Tam, J. S. (2021). Risk prediction of COVID-19 related death and hospital admission in adults after COVID-19 vaccination: National prospective cohort study. *BMJ*, *374*, n2244. https://doi.org/10.1136/bmj.n2244

Hughes, A., Barnard, S., Bauer-Staeb, C., Holleyman, R., Dunn, S., Fryers, P., Newton, J., Fitzpatrick, J., Burton, P., & Goldblatt, P. (2025). *Life lost due to COVID-19: analysis of mortality displacement in a National English Cohort*. https://doi.org/10.2139/ssrn.5097193. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5097193. Date accessed May 16, 2025.

Nafilyan, V., Humberstone, B., Mehta, N., Diamond, I., Coupland, C., Lorenzi, L., Pawelek, P., Schofield, R., Morgan, J., Brown, P., Lyons, R., Sheikh, A., & Hippisley-Cox, J. (2021). An external validation of the QCovid risk prediction algorithm for risk of mortality from COVID-19 in adults: A national validation cohort study in England. *The Lancet: Digital Health*, *3*(7), e425–e433. https://doi.org/10.1016/S2589-7500(21)00080-7

Torres-Collado, L., García de la Hera, M., Compañ-Gabucio, L. M., Oncina-Cánovas, A., González-Palacios, S., Notario-Barandiaran, L., & Vioque, J. (2022). Self-reported health status and mortality from all-causes of death, cardiovascular disease and cancer in an older adult population in Spain. *PLoS One*, *17*(1), e0261782. https://doi.org/10.1371/journal.pone.0261782

Walker, S., Goodfellow, H., Pookarnjanamorakot, P., Murray, E., Bindman, J., Blandford, A., Bradbury, K., Cooper, B., Hamilton, F. L., Hurst, J. R., Hylton, H., Linke, S., Pfeffer, P., Ricketts, W., Robson, C., Stevenson, F. A., Sunkersing, D., Wang, J., Gomes, M., Henley, W.; & Collaboration LWCR (2023). Impact of fatigue as the primary determinant of functional limitations among patients with

post-COVID-19 syndrome: A cross-sectional observational study. *BMJ Open*, *13*(6), e069217. https://doi.org/10.1136/bmjopen-2022-069217
WHO (2022). *International Statistical Classification of Diseases and Related Health Problems (ICD11)*. https://www.who.int/standards/classifications/classification-of-diseases. Date accessed May 16, 2025.

https://doi.org/10.1093/jrsssa/qnaf111
Advance access publication 7 July 2025



# Sheila M Bird's invited contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Sheila M. Bird**[1,2]

[1]Programme Leader Emerita, MRC Biostatistics Unit, Cambridge, UK
[2]Honorary Professor, College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK
*Address for correspondence*: Sheila M. Bird, MRC Biostatistics Unit, Robinson Way, Cambridge CB2 0SR, UK.
Email: sheila.bird@mrc-bsu.cam.ac.uk

My career as biostatistician spans five decades of epidemics, mainly batting for the Medical Research Council, but now with affiliation also to Edinburgh University. Wood et al. present useful international evidence on peoples' precautionary and distancing response ahead of lockdowns.

But the Royal Statistical Society (RSS) COVID-19 Taskforce was concerned that Test, Trace, and Isolate (TTI) was instructing close contacts of index cases to quarantine for 14 days *without learning efficiently about:*

(i) **how soon,** during their ***intended*** quarantine period, close contacts actually tested **PCR-positive for SARS-CoV-2:** be they (a) members of the household of an index case or (b) external contacts;
(ii) **adherence to quarantine;**
(iii) **influential covariates for transmission to close contacts; or for breaching quarantine.**

Hence, in July 2020, RSS COVID-19 Taskforce recommended two key statistical methods (https://rss.org.uk/RSS/media/File-library/Policy/RSS-COVID-19-Task-Force-Statement-on-TTI-final.pdf). First, that a running **random sample of households of types (a) and (b) be selected:** each to be ***visited on a pair of randomly sampled days*** during the ***intended*** quarantine period (e.g. first visit during days 1–5; second during days 6–14) to offer PCR-testing, observe adherence to quarantine, and record brief covariate information. Since TTI might be delayed in reaching close contacts, key information includes ***elapsed time from intended quarantine-start until the contact is reached by TTI.***

I suggested that these random visits could be delivered efficiently by the field-force of the Office for National Statistics Community Infection Survey. My suggestion was not adopted by Ministers. Instead, UK's duration of quarantine was shortened from 14 through 10 to 7 days on a non-experimental, eventual basis.

Next time, from the outset, UK should adopt RSS's two key statistical methods for gleaning intelligence from TTI-lookalikes. The second method, record-linkage to PCR-confirmed SAR-CoV-2 diagnoses in close contacts, shortly before, during or soon after quarantine, was adopted by October 2020; but putting the information into public domain was took longer.

Wood et al. suggest NICE's £30,000 per QALY as threshold for a nation's outlay when confronted by pandemic uncertainty. Early modelling alerted to likely 500,000 lives lost to SARS-CoV-2 in the UK. COVID-related deaths, mostly elderly, were counted as 10 life-years lost (not QALYs). As UK's actual toll was nearer 200,000, the national outlay 'saved' 3 million life-years (as central estimate) but at 10 times the NICE-affordable threshold.

Peace-time thresholds are not the correct reference for pandemics although an empirical multiplier thereof could be helpful. Wood et al. offer us a single UK multiplier [**10**] for SARS-CoV-2. Statisticians' hell is $N = 1$. Decision-makers need a distribution of multipliers: e.g. from other nations and other recent pandemics.

Unfortunately, the authors did not apply their 'reasoning' to HIV/AIDS; opioid-related deaths [which, in Scotland, claimed 1970 lives in 5 years (2006–2010) vs. 1864 lives lost to HIV/AIDS in 30 years (1983–2012)]; BSE/vCJD [fewer than 240 vCJD deaths, mainly in young people, internationally (Ritchie et al., 2021)]; and swine-flu (fewer 500 deaths in UK).

Pandemics, like wars, demand exceptional outlay: see ***The Long Shot*** about the UK Vaccines Taskforce in 2020 (Bingham & Hames, 2023).

The read paper challenges the presentation of risk and laments intentional distortion. The Bank of England's presentation of quantitative easing was egregious: a warning always to beware distortions of time axes; or their mis-labelling. As an aside, I recall in March 1996, when UK announced 10 vCJD cases in young people, the RSS President bracketed the likely toll as between 10 (cases to date) and 60 million (UK population) but was misreported internationally as having said between 10 million and 60 million.

During SARS-CoV-2, in early April 2021, presentation of a serious age-related adverse event risk, cerebral venous sinus thromboembolism with low platelet count (CVST), associated with the Oxford Astra Zeneca vaccine, was initially constrained in two respects. The young were at higher risk of CVST and UK initially decided that those under 30 years of age should no longer receive the Oxford AstraZeneca vaccine. Yet, no actual counts for Yellowcard reports of CVST by age-group or estimates of how many first doses were administered were shown; and no confidence intervals qualified the modelled CVST-risk by age-group. Moreover, the ultimate comparator in the UK was not 'unvaccinated' since UK citizens could receive messenger-RNA vaccines instead. Immediately, I advised my god-daughter, then in her 30s, to do just that. Within a month, the UK's at-unacceptable-CVST-risk threshold had shifted to include those aged 30–39 years.

The authors show that the impact of the post-war baby-boom on the calculation of expected deaths is notable and leads to error ***if the calculation is short-circuited.*** The ONS has improved the statistical underpinning of its work on expected deaths and, I hope, that ONS both eschews the emotive term 'excess deaths' and differentiates 'death-**occurrence**-week' from 'death-registration-week'. The two differ importantly.

Notably, the **UK's first COVID-related death occurred on 31 January 2020** but was not so-registered until September 2020 after the family had asked for their father's autopsy findings to be reviewed. Hence, 25th April 2020 was 85 days, or **12 weeks**, after the UK's first COVID-death (not **7 weeks**). Muddling death-**occurrence**-date and lagged death-registration-date catches us all out!

Latterly, the ONS not only reported COVID-related deaths by **occurrence**-week but also, by accounting for registration-delays, gave uncertainty estimates for the most recent weekly-death-**occurrences**.

Since swine-flu, the RSS has been calling for legislation to end the late registration of fact-of-death for coroner-referred and inquest-deaths in England, Wales, and Northern Ireland (Bird, 2013). In Scotland—for all deaths—fact-of-death is registered within 8 days of death having been ascertained. Hence, National Records for Scotland knows about deaths for whom cause has yet to be determined and can follow-up: ONS does not and cannot.

**President, I** appeal again: to the National Statistician, Chief Medical Officers and Chief Coroner, UK Statistics Authority, Government Chief Scientific Adviser (GCSA), Ministers & their CSAs to legislate for prompt registration of fact-of-death in England, Wales, and Northern Ireland—as in Scotland. And for this to be enacted within 18 months, that is: **before the new Health Data Research Service** comes into play.

By reference to locked-down vs. keyworker/nosocomial compartments, a neat explanation is given in Section 4.2 about the behaviour of the reproductive number, R which modelling needs

to accommodate: R is increasingly suppressed in the locked-down compartment and so, over time, R drifts upwards in the keyworker/nosocomial compartment(s).

Wood et al. judge harshly: 'Models were not validated for prediction in advance . . . that COVID spread could be massively reduced by suppressing human contact to the maximum extent possible is an especially undemanding check of model sanity'.

But I endorse strongly the need for a better balance between 'theory and data; modelling and measurement'. Not **any** data: but **data-by-design** to ensure analysability and inference. Data-by-design; efficient analysis; protocol; and analysts who had walked the walk (not consultants who had just talked the talk) were exemplified by the cross-sectional REACT studies. After all, statisticians can surely organize a party in a distillery!

And so, I am pleased to congratulate the authors on the wisdom 'distilled' in their provocative and thoughtful paper.

*Conflicts of interest:* none declared.

## References

Bingham, K., & Hames, T. (2023). *The long shot. The inside story of the race to vaccinate Britain*. Oneworld Publications.

Bird, S. M. (2013). Editorial: Counting the dead properly and promptly. *Journal of the Royal Statistics Society Series A*, *176*(4), 815–817. https://doi.org/10.1111/rssa.12035

Ritchie, D. L., Peden, A. H., & Barria, M. A. (2021). Variant CJD: Reflections a quarter of a century on. *Pathogens*, *10*(11), 1413. https://doi.org/10.3390/pathogens10111413

https://doi.org/10.1093/jrsssa/qnaf075
Advance access publication 26 June 2025

# Daniela De Angelis' invited contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al. with contributions from Joel Kandian and Paul Birrell

**Daniela De Angelis[1,2], Joel Kandiah[1] and Paul J. Birrell[1,2]**

[1]MRC Biostatistics Unit, University of Cambridge, East Forvie Building, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK
[2]Analysis and Intelligence Assessment, Chief Data Officer Group, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, UK

*Address for correspondence*: Daniela De Angelis, Medical Research Council Biostatistics Unit, East Forvie Building, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK. Email: daniela.deangelis@mrc-bsu.cam.ac.uk

This paper is very much a retrospective (and selective) journey through a unique situation in the history of the UK, when scientific evidence had to be produced under unprecedented pressure, with literally only hours to use messy and contradictory data to shed some light in a fog. We should bear in mind that decisions would have been made without any quantitative input otherwise. As statisticians, of course, we would have liked, as we teach our students, to carefully consider model choice and validation, and to carefully quantify uncertainty. This was a luxury that we could not afford then. In early 2020, the situation was unfolding rapidly, data from China were sparse and minimally

informative; and the pandemic in Italy was escalating with terrifying messages. We simply had to do our best given the information we had at the time. This is an important point that is not particularly acknowledged in this paper. Clearly, in retrospect, we can have a healthy and productive debate about statistical aspects. It is harder to have a productive debate about the conduct of politicians and the media, although my direct experience with the Science Media Centre has been excellent. This discussion will be limited to the data and consequent statistical and modelling aspects of this work.

## 1 Available data

### 1.1 Case data

I will start by agreeing with the authors about the data on COVID cases: it was an accounting exercise rather than a careful epidemiological data collection. If we had been carrying out a carefully designed data collection, information on symptom onset would have been collected, together with information on the reasons for testing. Such data would have let us disentangle the signal from the noise resulting from spatial and time-varying opportunistic testing as well as fluctuations due to test availability. Systematically collected data of this type are usually fantastically informative, particularly if we can correct for the inevitable delays in reporting (there exists a rich literature on methods for delay adjustments). This was not the case for the COVID case data, even with the delay correction. We tried to make sense of these data with no success. I would also agree that these are incidence, not prevalence, data.

### 1.2 Office of National Statistics Coronavirus Survey

The authors make an unfair comment on the Office of National Statistics Coronavirus Survey (ONS CIS). It took time to set up, as it was logistically very complex, but it was a unique resource to the UK of which other countries were envious. I would like to believe, like the authors, that availability of a CIS-type stream of data would address all the data requirements. However, as we are working on a reanalysis of the CIS, I can fully appreciate that this is not totally straightforward, with one immediate complication being its access through a secure research environment, which delays analysis and production of results. The authors talk about the need for publicly available data, but it is important to bear in mind that the ONS has an obligation to protect privacy.

## 2 Lockdowns

In this section of the paper, the authors state the 'retrospective belief that the lockdowns were necessary …seems to be based on informal reasoning and a priori modelling rather than data'. I have a number of comments. Firstly, I found this confusing as they call 'data' the information provided by retrospective estimates of 'current' incidence. This estimation is not straightforward. I wonder why the authors decided to estimate the 'fatal incidence', for example. Wood (2022) touches on this, but in this discussion paper no alternative estimates of incidence are provided. Estimates of the time-varying probability of dying in hospital have been available for a while (e.g. Kirwan et al., 2022) and could have been used to estimate incidence in the population, rather than incidence in those who die. This would, in practice, amount to scaling the deaths that Wood et al. use by the inverse of the probability of dying in hospital, given infection. This probability changed over time as improved treatment, together with an increasing fraction of people dying at home after the lockdown, reduced the probability of dying in hospital. Therefore, deaths would have been scaled up by a larger factor after the lockdown, which might have pushed the estimated peak in incidence to a later time. Again, Wood (2022) mentions this point, but he has chosen not to highlight it in the current discussion paper. This incidence estimate would have provided a fairer comparison with the 'incidence' from round 5 of the React2 survey (Ward et al., 2023), which is in itself only an approximation relying on the accurate recollection of the incidence of first symptoms (e.g. recall bias?) and the precision of the assumed delay between infection and symptoms (5.8 days). Secondly, trends in current infections were not the only quantity of relevance. More important was the impact of these infections on the health system, and that the National Health Service (NHS) should not be overwhelmed. This is an aspect the authors do not mention that we examine below and which is the basis of the contribution of Kevin Fong. Thirdly, though incidence estimates could be, in principle, used in a retrospective assessment, they were not available prospectively at the point where the decision on the lockdown needed to be made. At the time of the first lockdown, we only had

information on the number of deaths growing rapidly over time. What else, apart from trends in observed data, i.e. deaths or other measures of burden (e.g. hospitalizations), which are lagged and may even be affected by reporting delays, could inform decisions? The authors mention the CIS, but the CIS was not available at that stage, so only a synthesis of evidence from multiple indicators could inform decisions. Looking at the international comparisons: do these make sense, given differences in data definitions as well as differences in culture (e.g. Sweden) and health systems?

### *2.1 The alternative lockdown narrative*

For this discussion, it is useful to recall some aspects of the semi-stochastic real-time model (RTM; P. Birrell et al., 2021) we used during the first wave. In this model, the incidence of infection depends on the contact between susceptible and infectious and a parameter $\beta$ (the authors refer to it as 'modifying the contact rate'), which changes over time in weekly steps. The contact matrix and these parameters are, in the original model (OM), kept fixed until the lockdown is introduced. To address the criticism made that 'increasing incidence until the lockdown was simply built into the model', we extend the work in Kandiah et al. (2025) where we reanalyse data available at the time, investigating alternative model structures and the incorporation of additional information on contacts. Here, we compare estimates of transmission from the OM (P. Birrell et al., 2021) where the transmission parameter $\beta$ and the contact matrix are constant with estimates from two additional models: the Earlier Transmissibility Process(ETP) model, where $\beta$ is allowed to change before the lockdown, but the contact matrix is still kept constant; and the Earlier Contact Data (ECD) model (Kandiah et al., 2025), where both $\beta$ and the contact matrix change before the lockdown. This comparison, expressed in terms of $\mathcal{R}_t$, is shown in Figure 1.

Figure 1b shows that, when allowing $\beta$ to change before the lockdown, $\mathcal{R}_t$ does indeed decrease but remains well above 1, resulting in an infection incidence still increasing. This is confirmed when we also allow the contact matrix to change to include information on behaviour prelockdown (Figure 1c). Only after the lockdown we estimate $\mathcal{R}_t < 1$, with the number of infections decreasing. This is clearly shown in Figure 2 where incidence estimates from the OM, ETP, and ECD models are compared with a scaled version of the fatal incidence estimate by the authors.

Why is there such a difference? Wood's model and the RTM of Kandiah et al. (2025) and P. J. Birrell et al. (2025), are, at their heart, deconvolution approaches. Using Wood's convolution equation

$$\mathbb{E}\big[\text{deaths on day } t_i\big] = \mu(t_i) = \sum_{d=0}^{D_i} \exp\big\{ f\big(t_i - d\big)\big\}\pi\big(d\big),$$

the two approaches provide very different specifications for $f(\cdot)$. The RTM also includes serological information, though these data are not highly influential in the first phase of the pandemic, due to low and uncertain sensitivity of the early assay and waning immunological response.

One of the most immediate aspects to investigate is the difference in the infection-to-death distribution used in the two approaches. To this end, we modify the OM, EPT, and ECD models to use the same log-normal infection-to-death distribution used by Wood et al. The resulting estimated (national) incidence (Figure 3) shows infections increasing at the time of the lockdown, though there is a sharp decline in $\mathcal{R}_t$, which remains above 1. In the OM, the prelockdown decline in $\mathcal{R}_t$ can only be due to depletion of susceptible individuals. The increase immediately prior to the lockdown results from the diminishing contribution of London to the overall $\mathcal{R}_t$, which, notably, is an average over *infections*, not the whole population. This is not a realistic result, but motivates a deeper look at the regionally stratified results (Figure 4), which, across the different model variations, show that in London $\mathcal{R}_t < 1$ prior to the lockdown. This behaviour for $\mathcal{R}_t$ is not common across regions other than London, which closely resemble the North-East and Yorkshire shown here. The extent to which London is driving the estimated infection dynamics of the first wave is further explored in the contribution of Paul Birrell.

Apart from reducing transmission and the number of new infections, there was a pressing need to avoid overwhelming the NHS. As in , it is possible to generate the number of deaths averted by the lockdown by assuming a counterfactual scenario in which the lockdown is not imposed, and

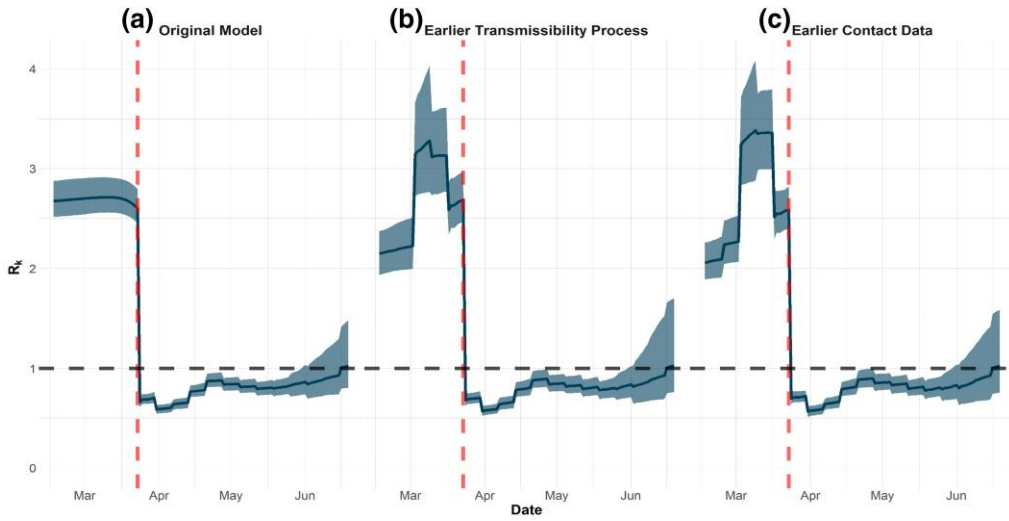



**Figure 1.** Posterior median and 95% Credible Interval for $\mathcal{R}_t$ for England by model.

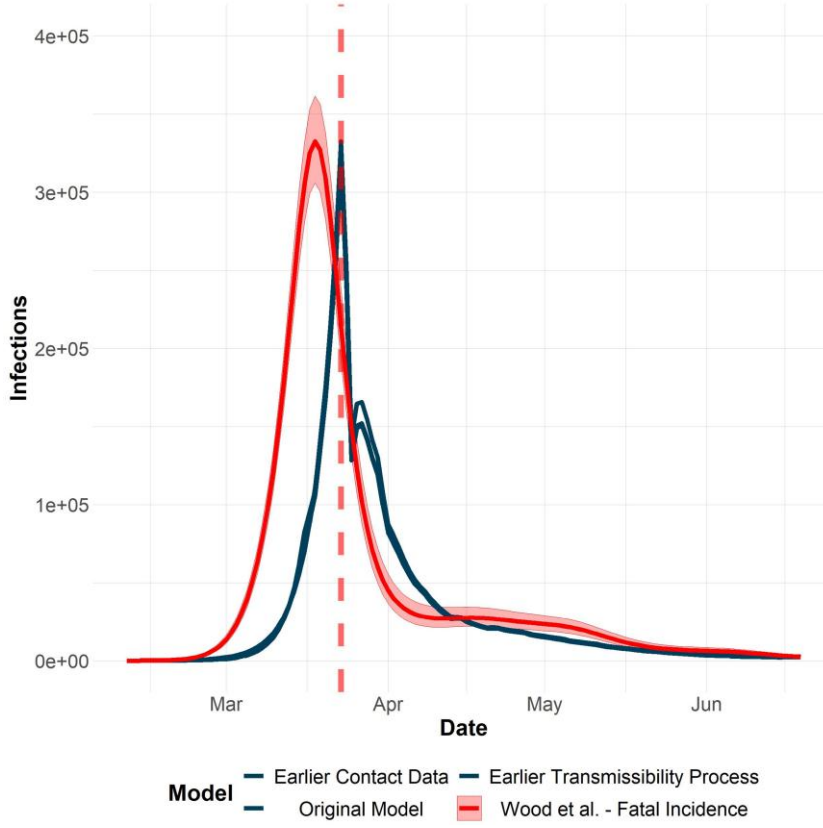



**Figure 2.** Comparison between estimates of incidence from the original model (OM), Earlier Transmissibility Process (ETP), and Earlier Contact Data (ECD) models and the fatal incidence in this paper.

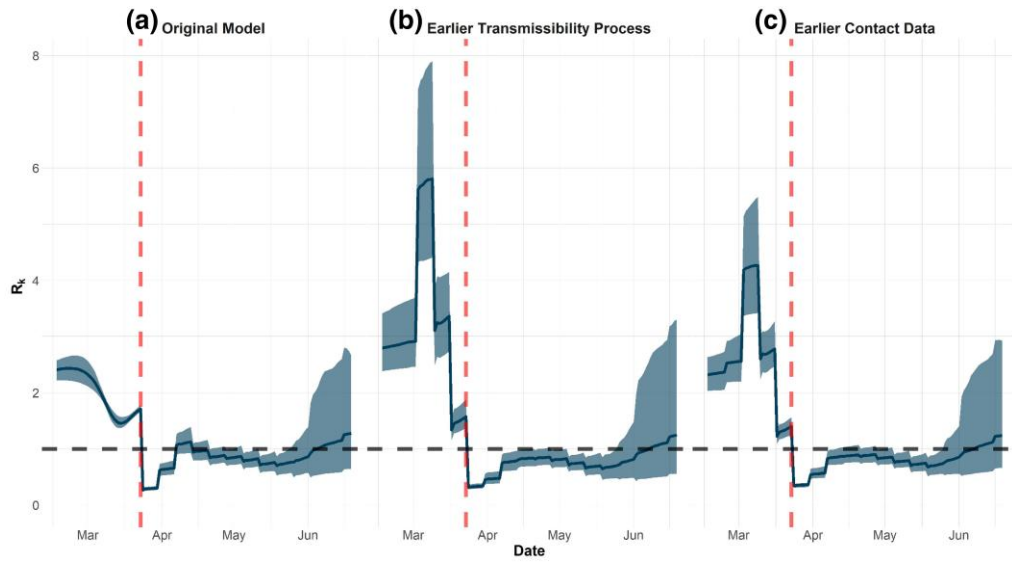



**Figure 3.** Posterior median and 95% Credible Interval for $\mathcal{R}_t$ for England by model, when using a log-normal infection-to-death distribution, as in Wood et al.

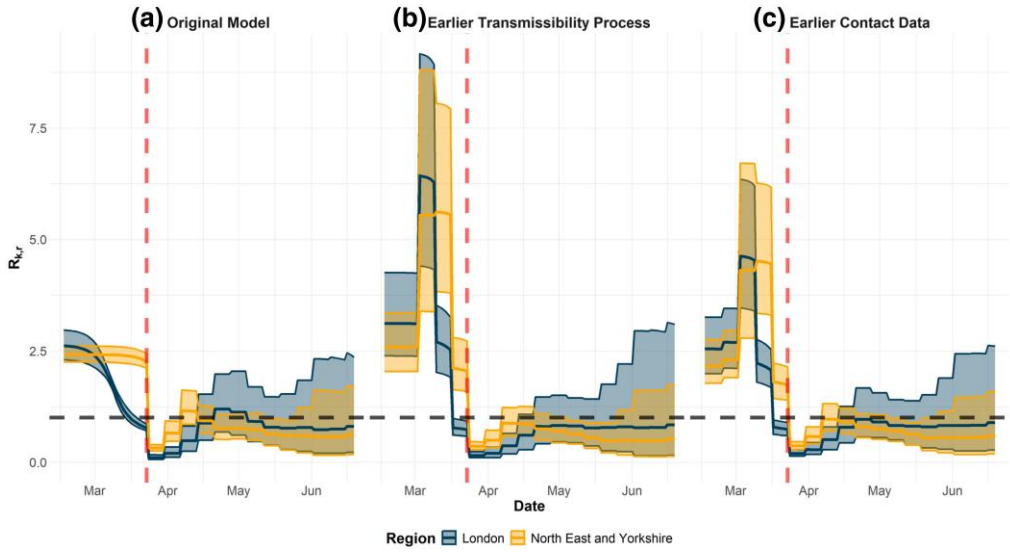



**Figure 4.** Posterior median and 95% Credible Interval for $\mathcal{R}_t$ for England by model and region, when using a log-normal infection-to-death distribution, as in Wood et al.

all transmission parameters and contact matrices remain constant at their values immediately prior to the lockdown (see P. J. Birrell et al., 2025). Even in London, where a prelockdown peak in incidence has been estimated with certainty, nearly 3,000 deaths are averted under the lockdown (see Table 1), despite the already declining incidence. This alerts to a significant additional and continuing burden, with which healthcare services may not have been able to cope (see the contribution of Kevin Fong).

I want to conclude by making a few remarks on claims in the paper and in Simon Wood's statements during an interview about this paper: it was not the case, in my experience, that 'the Government followed the part of the science that confirmed that the policies that had been

**Table 1.** Deaths averted by the lockdown, assuming RTM transmission parameters and mobility remain constant

| Region | Deaths averted |
|---|---|
| England | 178,000 (124,000–236,000) |
| East of England | 15,100 (4,600–31,600) |
| London | 2,700 (1,090–5,610) |
| Midlands | 10,400 (2,420–27,600) |
| North-East and Yorkshire | 45,700 (23,400–66,100) |
| North-West | 45,000 (30,700–62,100) |
| South-East | 28,200 (9,880–54,600) |
| South-West | 29,000 (8,110–52,000) |

RTM≙ real-time model.

implemented had been right'. There was a very open process of debate and interminable discussions before a consensus would be reached [at least in the Scientific Pandemic Influenza Advisory Committee, Subgroup on Modelling (SPI-M)]. It is not the case 'that science was overly dominated by epidemic models'. These models are the most appropriate tools in outbreak situations. This led to having more modellers involved in the work, as statistical expertise in this area was lacking.

I feel it is true, however, that more statisticians should have been involved in the pandemic response and that statisticians should be more assertive about the data skills they bring. To this end, the Royal Statistical Society has launched the *Statistics Under Pressure Steering Group*. I hope this will serve to educate and support statisticians on how to work effectively in emergency situations, so that more can participate in groups like SPI-M. In fact, I myself suggested the creation of a SPI-Statistics subgroup, which I very much hope will become a reality.

I would like to thank the authors for their highly provocative paper and the Royal Statistical Society for giving me the opportunity to discuss it.

*Conflicts of interest:* D.D.A and P.J.B. are members of the Scientific Pandemic Infections group on Modelling ("SPI-M"), an advisory group to the UK government's Department of Health and Social Care (DHSC), designed to provide expert advice based on infectious disease modelling. The content of this response reflects personal opinions and not those of DHSC or UKHSA.

## References

Birrell P., Blake J., van Leeuwen E., Gent N., & De Angelis D. (2021, May). Real-time nowcasting and forecasting of COVID-19 dynamics in England: The first wave. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200279. https://doi.org/10.1098/rstb.2020.0279

Birrell P. J., Blake J., Kandiah J., Alexopoulos A., van Leeuwen E., Pouwels K. B., Ghosh S., Starr C., Walker A. S., House T. A., Gay N., Finnie T., Gent N., Charlett A., & De Angelis D. (2025, April). Real-time modelling of the SARS-CoV-2 pandemic in England 2020–2023: A challenging data integration. *Journal of the Royal Statistical Society Series A: Statistics in Society*, qnaf030. https://doi.org/10.1093/jrsssa/qnaf030

Kandiah J., van Leeuwen E., Birrell P. J., & De Angelis D. (2025, August). Contact data and SARS-CoV-2: Retrospective analysis of the estimated impact of the first UK lockdown. *J Theor Biol*, *610*, 112158. https://doi.org/10.1016/j.jtbi.2025.112158

Kirwan P. D., Charlett A., Birrell P., Elgohari S., Hope R., Mandal S., De Angelis D., & Presanis A. M. (2022, August). Trends in covid-19 hospital outcomes in England before and after vaccine introduction, a cohort study. *Nature Communications*, *13*(1), 1–11. https://doi.org/10.1038/s41467-022-32458-y

Ward H., Atchison C., Whitaker M., Davies B., Ashby D., Darzi A., Chadeau-Hyam M., Riley S., Donnelly C., Barclay W., Cooke G., & Elliott P. (2023, November). Design and implementation of a national program to monitor the prevalence of SARS-CoV-2 IGG antibodies in England using self-testing: The react-2 study. *American Journal of Public Health*, *113*(11), 1201–1209. https://doi.org/10.2105/AJPH.2023.307381

Wood S. N. (2022, September). Inferring UK covid-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? *Biometrics*, *78*(3), 1127–1140. https://doi.org/10.1111/biom.v78.3

**The vote of thanks was passed by acclamation**

https://doi.org/10.1093/jrsssa/qnaf150
Advance access publication 10 December 2025



# M. Gabriela M. Gomes, Ibrahim Mohammed and Chris Robertson's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**M. Gabriela M. Gomes[1,2], Ibrahim Mohammed[1,3] and Chris Robertson[1,4]**

[1]Department of Mathematics and Statistics, University of Strathclyde, Glasgow G1 1XH, UK
[2]Centre for Mathematics and Applications (NOVA MATH), NOVA School of Science and Technology, Caparica 2829-516, Portugal
[3]Department of Mathematical Sciences, Abubakar Tafawa Balewa University, Bauchi PMB 0248, Nigeria
[4]Public Health Scotland, Glasgow G2 6QE, UK

*Address for correspondence*: M. Gabriela M. Gomes, Department of Mathematics and Statistics, University of Strathclyde, Glasgow, G1 1XH, UK. Email: gabriela.gomes@strath.ac.uk

In Section 4 of their paper (Wood et al., 2025), the authors address the omission of person-to-person variability in Covid-19 transmission models selected to inform policy (such as SEIR models). Modelling efforts incorporating unobserved individual variation in susceptibility and exposure to infection (Gomes et al., 2022, akin frailty models established in survival analysis (e.g. Balan & Putter, 2020), were excluded from policy research on the basis of supposed parameter identifiability issues. This contribution aims to refute the view that the frailty parameter is especially accountable for lack of identifiability in SEIR models.

For illustration, use the SEIR model with individual variation in susceptibility (Gomes et al., 2022; Wood et al., 2025)

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -c(t)R_0\gamma IS^{\lambda}, \quad \frac{\mathrm{d}E}{\mathrm{d}t} = c(t)R_0\gamma IS^{\lambda} - \delta E, \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \delta E - \gamma I,$$

where $S$, $E$, and $I$ are susceptible, exposed, and infectious population fractions (the remaining fraction $R$ being recovered or removed), $\delta = 0.18$ per day, $\gamma = 0.25$ per day, $\lambda = 1 + v^2$, where $v$ is the coefficient of variation of individual susceptibility, $R_0$ is the basic reproduction number, and $c(t)$ is a time-dependent multiplicative factor representing voluntary behaviour changes and compliance with NPIs (starting at 1 and, from $t = t_0$ onward, decreasing towards $c_1$ (Figure 1, bottom)). The model with $R_0 = 3$, $\lambda = 3$, $t_0 = 15$ days, and $c_1 = 0.3$, was run forward in time starting from two alternative sets of initial conditions to generate the coloured trajectories in Figure 1 (middle). For comparison, the scenario where $c(t) = 1$ is also included in Figure 1 (top).

To assess parameter identifiability, sets of synthetic data were generated with a stochastic version of the model and parameters $R_0$, $v$, $t_0$, and $c_1$ were estimated by maximum likelihood. Simulated scenarios assumed $R_0 = 3$, $v = \sqrt{2}$, $t_0 = 15$ days, and $c_1 = 0.2$ or 0.3.

First, we fitted the model to data from a single epidemic curve generated with initial infectious fraction $I_1(0) = 2.4e-4$. The procedure was conducted 200 times and correlations between

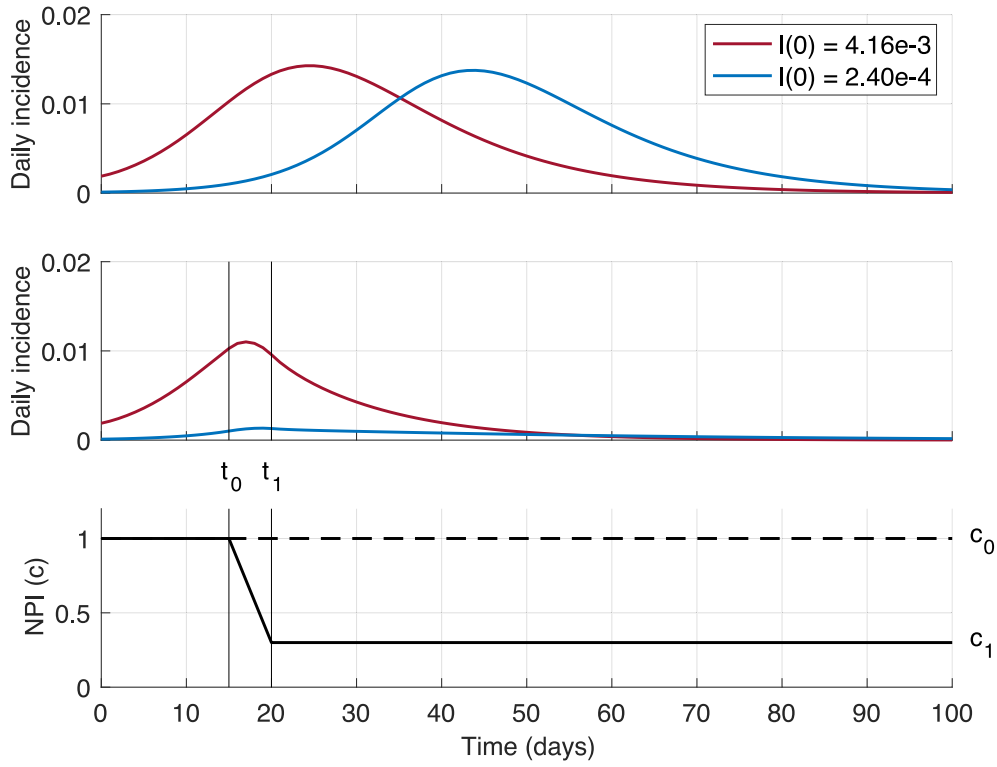



**Figure 1.** Incidence trajectories generated by the SEIR model with individual variation in susceptibility. The two curves on top and middle panels correspond to different initial conditions (red ~ 17 times higher prevalence than blue when $t = 0$). Bottom panel is the multiplicative factor for transmission due to voluntary behaviour change and NPI.

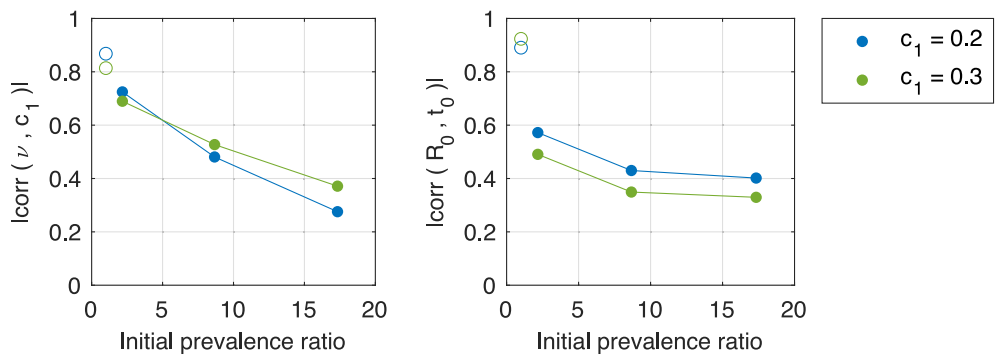



**Figure 2.** Parameter correlations in the SEIR model. Pairwise correlations obtained from the Hessian of MLEs based on simulated data of a single epidemic (open circles) and two concurrent epidemics which differ in initial infectious prevalence (dots). The left panel accounts for individual variation in susceptibility represented by parameter $\lambda$ (or $\nu$) while the right panel assumes a homogeneous population by imposing $\lambda = 1$ (or $\nu = 0$).

parameters were obtained from the Hessian matrix of the maximum likelihood estimates (MLEs). We found strong positive correlations between $\nu$ and $c_1$: 0.87 when $c_1 = 0.2$, and 0.81 when $c_1 = 0.3$ (open circles in Figure 2 (left)). This is not surprising as both increasing $\nu$ and decreasing $c_1$ flatten the curve.

Second, we fitted the model to two concurrent epidemics (which might be occurring in different regions) assuming that the same parameters applied to both but initial conditions were different. The approach resulted in a marked reduction in $|\mathrm{Corr}(\nu, c_1)|$ (dots in Figure 2 (left)). This outlines

an inference approach, originally applied to Covid-19 in England and Scotland (Gomes et al., 2022), to overcome common challenges in the inference of unobserved individual variation from population data (Balan & Putter, 2020; Gomes et al., 2024).

Finally, we conducted a similar analysis with a version of a model that assumes no individual variation ($\lambda = 1$). In the single epidemic analysis, we identified strong negative correlations between $R_0$ and $t_0$: $-0.89$ when $c_1 = 0.2$, and $-0.92$ when $c_1 = 0.3$ (open circles in Figure 2 (right)). Introducing a second epidemic in the analysis alleviated $|\text{Corr}(R_0, t_0)|$ (dots in Figure 2 (right)) much like in the scenario of $|\text{Corr}(v, c_1)|$ above. Therefore, the identifiability issues in SEIR models do not originate from the introduction of $v$.

*Conflicts of interest:* None declared.

## Funding

M.G.M.G. is partly funded by Fundação para a Ciência e a Tecnologia, I.P., under the scope of projects UIDB/00297/2020 (https://doi.org/10.54499/UIDB/00297/2020) and UIDP/00297/2020 (https://doi.org/10.54499/UIDP/00297/2020).

## References

Balan T. A., & Putter H. A. (2020). A tutorial on frailty models. *Statistical Methods in Medical Research*, *29*(11), 3424–3454. https://doi.org/10.1177/0962280220921889

Gomes M. G. M., Blagborough A. M., Langwig K. E., & Ringwald B. (2024). Remodelling selection to optimise disease forecasts and policies. *Journal of Physics: A, Mathematical and Theoretical*, *57*(10), 103001. https://doi.org/10.1088/1751-8121/ad280d

Gomes M. G. M., Ferreira M. U., Corder R. M., King J. G., Souto-Maior C., Penha-Gonçalves C., Gonçalves G., Chikina M., Pegden W., & Aguas R. (2022). Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *Journal of Theoretical Biology*, *540*, 111063. https://doi.org/10.1016/j.jtbi.2022.111063

Wood S. N., Wit E. C., McKeigue P. M., Hu D., Flood B., Corcoran L., & Jawad T. A. (2025). Some statistical aspects of the covid-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society (to appear)*.

https://doi.org/10.1093/jrsssa/qnaf104
Advance access publication 2 July 2025

# Irons' contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Nicholas J. Irons**

Department of Statistics, Leverhulme Centre for Demographic Science, and Pandemic Sciences Institute, University of Oxford, Oxford OX1 3LB, UK

*Address for correspondence*: Nicholas J. Irons, Department of Statistics, 24-29 St Giles', Oxford OX1 3LB, UK. Email: nicholas.irons@stats.ox.ac.uk

I thank Wood et al. (2025) for raising a discussion about the COVID response. Here, I focus on their points about the effectiveness and cost-effectiveness of nonpharmaceutical interventions (NPIs), which I have studied (Irons & Raftery, 2024).

Regarding cost-effectiveness, the authors are right to point out the social, health, and economic trade-offs incurred by restrictions. However, they neglect the economic disruption caused by the pandemic itself, even in the absence of interventions. This is a significant omission. We know that people

voluntarily reduced their own economic activity in response to the threat of virus. Furthermore, worker and student absenteeism resulting from the viral spread led to the closures of businesses and schools. The absence of interventions is not business as usual; that is not the relevant counterfactual.

As a result, the authors do not make a fair comparison of COVID and economic outcomes. They inappropriately attribute all extra borrowing and lost economic activity to interventions. In the U.S., Gupta et al. (2023) found that only 60% of the 12 pp decline in the employment rate in spring 2020 could be explained by government policies. Similarly, Bodenstein et al. (2022) found that the combined effect of voluntary and mandatory social distancing could explain only one-half to two-thirds of the 12% drop in GDP in the second quarter of 2020.

Turning to the effectiveness of interventions in reducing transmission, I cannot argue with the authors' analysis of the European response, as I have not analysed the data myself nor attempted to reproduce their results. (Although, I am curious of the effect of all that smoothing on downstream inferences about transmission rates.) From the US response, however, it is clear that infections did not generally peak prior to the onset of restrictions. Indeed, Irons and Raftery (2024) found that NPIs were effective in reducing transmission, saving an estimated 860,000 lives in the U.S. in 2020 alone. We also found that masking, testing, tracing, social distancing orders, and reactive workplace closures were all cost-effective measures, while school closure was not.

I agree with Wood et al. (2025) that a careful review of pandemic decision-making is urgently needed. Such assessments must be based on a rigorous and honest evaluation of the evidence, carefully estimating and weighing the benefits and costs of interventions. I hope that, for our own sake, we can come to a consensus on this important question.

*Conflicts of interest:* None declared.

## References

Bodenstein M., Corsetti G., & Guerrieri L. (2022). Social distancing and supply disruptions in a pandemic. *Quantitative Economics*, *13*(2), 681–721. https://doi.org/10.3982/QE1618

Gupta S., Montenovo L., Nguyen T., Lozano-Rojas F., Schmutte I., Simon K., Weinberg B. A., & Wing C. (2023). Effects of social distancing policy on labor market outcomes. *Contemporary Economic Policy*, *41*(1), 166–193. https://doi.org/10.1111/coep.12582

Irons N. J., & Raftery A. E. (2024). *US COVID-19 school closure was not cost-effective, but other measures were*. arXiv:2411.12016. https://doi.org/10.48550/arXiv.2411.12016, preprint: not peer reviewed.

Wood S. N., Wit E. C., McKeigue P. M., Hu D., Flood B., Corcoran L., & Abou Jawad T. (2026). Some statistical aspects of the Covid-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *189*(1), 1–31. https://doi.org/10.1093/jrsssa/qnaf092

https://doi.org/10.1093/jrsssa/qnaf083
Advance access publication 27 June 2025

# Nicholas Fisher, Noel Cressie and Dennis Trewin's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Nicholas Fisher[1], Dennis Trewin[2] and Noel Cressie[3]**

[1]School of Mathematics & Statistics, University of Sydney, Sydney, Australia
[2]Faculty of Health, University of Canberra, Canberra, Australia
[3]School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, Australia

*Address for correspondence:* Dennis Trewin, Faculty of Health, University of Canberra, Canberra, ACT 2617, Australia. Email: dennistrewin7@gmail.com

**Keywords:** Covid-19, sample surveys, data scientist

John Chambers (1993) made an important distinction between what he termed 'greater statistics' and 'lesser statistics'. In our view, the authors' article falls in between. While it is clearly addressing an important issue, its focus is *statistical modelling* of Covid-19 and is retrospective in nature. A broader forward-looking discussion is needed to prepare for the next pandemic.

The need for expeditious action in response to the emergence of Covid-19 was *already* evident at the beginning of 2020. As statistical/data scientists, we were urging the Australian government to start an ongoing national survey to capture data about the progress of the virus, based on sampling-design principles. Supportive interest came from some of the major media outlets with an op ed published, and a *Pandemic Information Management Plan* was submitted to our Department of Prime Minister and Cabinet in August 2020 (Fisher & Trewin, 2021). The plan outlined the various forms of data that should be acquired, the need to convert the data into actionable monitoring information, and the necessary collective skills, knowledge, and know-how.

Much like the article under discussion, the submission was critical of the undue emphasis on case data and the inherent biases that go with it.

In fact, an alert government could have developed and implemented such a management plan within the first half of 2020. A Chief Scientist, or ideally a *Chief Data Scientist*, would be able to take a strategic view of data needs and have appropriately high-level access to government leaders. That is 'greater statistics'. Many lives may well have been saved with expeditious action guided by such a management plan.

Australia and all forward-looking countries need to establish an Office of the Chief Data Scientist (OCDS)! The Covid-19 pandemic, the UK Post Office disaster (Bird & Howarth, 2024), the Australian Robodebt tragedy (Trewin *et al.*, 2023), and many other examples could have been substantially mitigated if not wholly avoided, were such an Office in place. More recently, a critical task of an OCDS would be to support government management of risks associated with artificial intelligence.

Why do these seemingly obvious and important recommendations keep getting ignored? Clearly, there is a lack of understanding among policy makers of the importance of uncertainty quantification and statistical thinking in making wise policy. To redress this, we should work harder on effective communication of our science of uncertainty to all parties at the policy-making table.

*Conflicts of interest:* None declared.

## References

Bird, S., & Howarth, M. (2024, May). Post Office scandal: Prosecutions data should have been a red flag. *Significance, Blog Post*, 2024.

Chambers, J. M. (1993). Greater or lesser statistics: A choice for future research. *Statistics and Computing*, *3*(4), 182–184. https://doi.org/10.1007/BF00141776

Fisher, N. I., & Trewin, D. J. (2021). A proposal to enhance national capability to manage epidemics: The critical importance of expert statistical input including official statistics. *Statistical Journal of the IAOS*, *37*(2), 465–481. https://doi.org/10.3233/SJI-210808

Trewin, D., Fisher, N., & Cressie, N. (2023). The Robodebt tragedy. *Significance*, *20*(6), 18–21. https://doi.org/10.1093/jrssig/qmad092

https://doi.org/10.1093/jrsssa/qnaf113
Advance access publication 4 July 2025

# Pieter Streicher and Alex Broadbent's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Pieter Streicher**

Centre for Philosophy of Epidemiology, Medicine and Public Health, University of Johannesburg, Johannesburg, South Africa

*Address for correspondence*: Pieter Streicher (PhD Eng), Research Associate, Centre for Philosophy of Epidemiology, Medicine and Public Health, University of Johannesburg, PO Box 524, Auckland Park, 2006 Johannesburg, South Africa. Email: pstreicher@uj.ac.za



This excellent and comprehensive paper highlights how statistical aspects of the COVID-19 response could have been improved. Critical reflection after a major crisis is essential, yet little has been done to evaluate the pandemic response—particularly the use of counterfactual disease modelling. Wood et al. identify significant errors and biases in the models that informed key strategies.

Pandemic interventions were often justified using counterfactual models predicting dire outcomes in the absence of action. However, these models were rarely evaluated retrospectively to test their adequacy. Instead, they were used primarily to justify stringent policies like lockdowns. Wood et al.'s Figure 15 shows that in cases where recommended interventions were not implemented—effectively turning the counterfactual into reality—the model predictions proved grossly inaccurate and unfit for policy guidance.

The paper also critiques how intervention timing was assessed using changes in the reproduction number ($R_t$). In many instances, $R_t$ fell below 1.0 before hard lockdowns were implemented, meaning the decline could not have been caused by the lockdown itself, as cause must precede effect.

Wood et al.'s credibility is enhanced by the fact that this retrospective evaluation was conducted early in 2020 using sound statistical methods, proving such analysis was both feasible and timely. Unfortunately, the peer review process was fraught with delays—possibly due to the paper's unpalatable conclusions rather than methodological flaws. As a result, crucial insights were not available in time to influence policy.

Notably, subsequent lockdown assessment papers by Imperial College London (Flaxman et al. 2020, Mishra et al. 2021, Knock et al. 2021) progressively increased the extent that $R_t$ dropped prior to the lockdown date. In the Flaxman paper, the drop in $R_t$ prior to lockdown was insignificant (see Appendix). In the Knock paper, $R_t$ had already declined significantly prior to lockdown and in London, $R_t$ was already well below 1.0 at the time of lockdown (see Appendix). The final Knock paper vindicated Wood et al.'s earlier findings even though some differences remain.

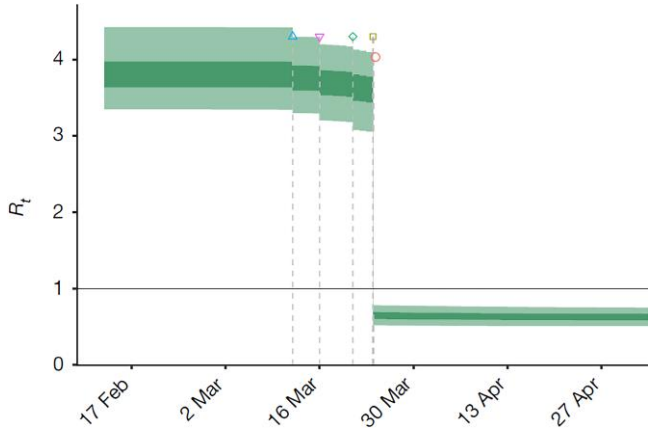
Our own research published in *Global Epidemiology* similarly argues that both retrospective evaluation and contemporaneous statistical sense-checking were lacking in the UK's SAGE-SPI-M modelling. This neglect led to repeated failures of mechanistic disease models. In contrast, teams like South Africa's SACMC, which later incorporated robust statistical methods alongside mechanistic ones, showed marked improvement over time.

[Link to article: https://www.sciencedirect.com/science/article/pii/S2590113324000439?via=ihub]

*Conflicts of interest:* none declared.

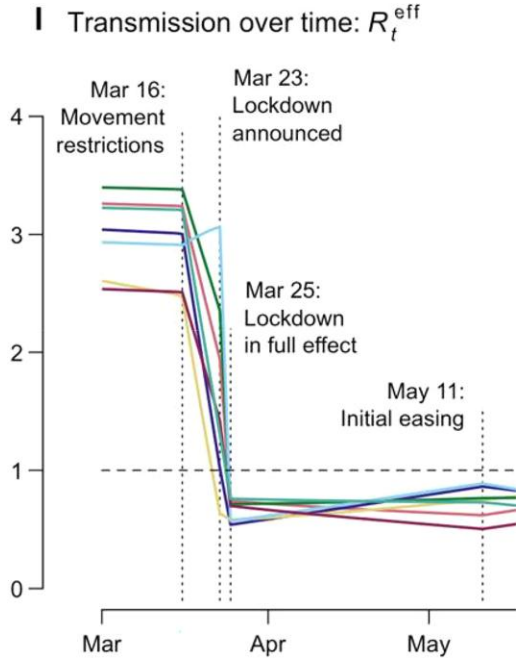
## Appendix

**ICL paper published on 8 June 2020:** Flaxman et al. *Estimating the effects of nonpharmaceutical interventions on COVID-19 in Europe*. https://doi.org/10.1038/s41586-020-2405-7




$R_t$ was high at around 4 in the UK, and then dropped abruptly to 0.7 coinciding exactly with the hard lockdown implemented on 23 March 2020. The drop in $R_t$ prior to the lockdown is insignificant.

**ICL paper published on 14 July 2021**. Knock et al. *Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England*. https://doi.org/10.1126/scitranslmed.abg4262




$R_t$ dropped significantly from just before 16 March already reaching 0.7 on the day that a full lockdown was announced in London (yellow line) on 23 March.

https://doi.org/10.1093/jrsssa/qnaf088
Advance access publication 18 July 2025

# Ursula Berger, Göran Kauermann, and Helmut Küchenhoff's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Ursula Berger, Göran Kauermann and Helmut Küchenhoff**

Department of Statistics, Ludwig-Maximilians-Universität München, Germany

*Address for correspondence*: Göran Kauermann, Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstrasse 33, 80539 Munuch, Germany.
Email: goeran.kauermann@lmu.de



First, we want to thank the authors for their persistence and elaborateness in putting this paper together. We honour the effort of our colleagues to look back and work through the COVID-19 pandemic with a critical but statistically scientific view. We admire and appreciate this, especially since in Germany, the last government had promised a review of the measures taken during the pandemic, but an initiative in this direction was never actually brought on the way. We want to add some aspects which touch on our own statistical experiences during those times.

First, Section 3 in the paper mirrors our own experiences. News and newspapers were filled with COVID-19, and numerous 'experts' found their platform—the more sensational the headline, the better. Apparently, with the dynamics of the pandemic, regular scientific publications were too slow. Therefore, we decided to release biweekly analysis reports, describing the pandemic situation statistically. Our reports—almost 30 in total—were internally reviewed, and we refrained from any speculation or political opinions. We showed that children were not pandemic drivers (Berger et al., 2022), that the peak number of infections in the first wave in Germany was before the first lockdown (Küchenhoff et al., 2021), Our findings often contradicted the prevailing media narrative, as exemplified below, see also Fritz et al. (2023).

Our second comment refers to excess mortality. We agree that adjustment for demographic changes is essential, as pointed out in Chapter 2.4 of the discussed paper. We performed calculations for excess mortality for 25 high-income countries and found a highly relevant effect of age adjustment (De Nicola & Kauermann, 2025). Interestingly, our result for the UK (72,000 excess deaths in 2020 and 2021) is nearly identical to the result in Figure 4 of the discussed paper. For Germany, excess mortality was low based on our calculations (De Niocla et al., 2022). Still, the Federal Statistical Office in Germany published a press release stating an excess mortality of 4.9% in 2020[1] compared with 2019. A closer look reveals that (a) 2019 had no flu wave, leading to 2% fewer deaths in 2019, (b) demographic aging accounted for an expected 2% increase in deaths in 2020, and (c) 2020 was a leap year, adding another 0.3% to mortality. Considering these factors, the 4.9% excess included 4.3%, which can be explained differently. Overall, we find it quite astonishing that there is no gold standard for age adjustments and that in some publications comparing excess mortality internationally, such adjustments were not applied at all (De Nicola & Kauermann, 2025).

*Conflict of interest:* none declared.

## References

Berger, U., Fritz, C., & Kauermann, G. (2022). Reihentestungen an Schulen können die Dunkelziffer von COVID-19 Infektionen unter Schülern signifikant senken. *Das Gesundheitswesen*, *84*(6), 495–502. https://doi.org/10.1055/a-1813-9778

De Nicola, G., & Kauermann, G. (2025). Estimating excess mortality in high-income countries during the COVID-19 pandemic. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *188*(1), 205–222. https://doi.org/10.1093/jrsssa/qnae031

[1] see: https://www.destatis.de/DE/Presse/Pressemitteilungen/2021/11/PD21_505_23211.html

De Nicola, G., Kauermann, G., & Höhle, M. (2022). On assessing excess mortality in Germany during the COVID-19 pandemic. *AStA Wirtschafts-und Sozialstatistisches Archiv*, *16*(1), 5–20. https://doi.org/10.1007/s11943-021-00297-w

Fritz, C., De Nicola, G., Günther, F., Rügamer, D., Rave, M., Schneble, M., Bender A, Weigert M, Brinks R, Hoyer A, Berger U, Küchenhoff H, & Kauermann, G. (2023). Challenges in interpreting epidemiological surveillance data–experiences from Germany. *Journal of Computational and Graphical Statistics*, *32*(3), 765–766. https://doi.org/10.1080/10618600.2022.2126482

Küchenhoff, H., Günther, F., Höhle, M., & Bender, A. (2021). Analysis of the early COVID-19 epidemic curve in Germany by regression models with change points. *Epidemiology & Infection*, *149*, e68. https://doi.org/10.1017/S0950268821000558

https://doi.org/10.1093/jrsssa/qnaf085
Advance access publication 30 June 2025



# Nick Andrews' contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Nick Andrews**

Immunisation and Vaccine Preventable Diseases Department, UK Health Security Agency, London, UK

*Address for correspondence*: Nick Andrews, Immunisation and Vaccine Preventable Diseases Department, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, UK. Email: nick.andrews@ukhsa.gov.uk

**Keywords:** Covid-19, excess mortality, pandemic

In the section on excess mortality, the authors use life tables based on 2017–2019 deaths to give an excess death estimate of 95,000 for the UK from 2020 to 2022. This is much lower than estimates from ONS and PHE where it was over 160,000. The PHE estimate of about 163,000 was obtained from what is now an Office of Health Improvement and Disparities (OHID) dashboard (GOV.UK, 2025).

The authors criticize the ONS method as it has a baseline based on the average mortality from 2015 to 2019 with no allowance for an aging population. Whilst this is a fair criticism, the same does not apply to the OHID method which also uses a 2015–2019 baseline, but does standardize for age, and furthermore includes a trend to allow for improving life expectancy.

Interestingly, an age standardized mortality approach was taken in two unreferenced reports produced by the Institute and Faculty of Actuaries in their Continuous Mortality Investigation (CMI) mortality monitor (Continuous Mortality Investigation, 2024) and in a paper comparing international excess mortality (WP180) (Continuous Mortality Investigation, 2025). In the first report, they use just deaths in 2019 as the comparator year and get an excess for 2020–2022 of 151,700. In the second, they use trends from 2010 to 2019 to give expected 2019 mortality which they then use as the comparator which gives a % excess that would translate to a similar excess mortality as in the first report.

So, why the discrepancy to the 95,000? This is very likely to be down to the particular baseline years used by Wood et al. and the failure to allow for continued life expectancy improvement. 2019 was a low mortality year whereas 2017 and 2018 were higher by about 4%. Averaging these with no trend for an improvement compared to just using 2019 or using more years with a trend means the baseline is about 2.7% higher, which over two years could account for about 40,000 deaths. Furthermore, starting from Week 1 rather than Week 10 2020 could explain 10,000 more as early 2020 was a low mortality period. The 10,000 difference between OHID and CMI is likely because OHID extrapolated life expectancy improvement into 2020–2022.

In summary, estimating excess over such a long period is difficult and very dependent on assumptions. Not assuming a continued life expectancy improvement along with the choice of baseline years for the life table has given a much lower excess in this paper compared to other more reasonable methods.

*Conflicts of interest:* None declared.

## References

Continuous Mortality Investigation. (2024). Mortality Monitor, Week 1 of 2024. https://www.actuaries.org.uk/documents/england-wales-mortality-monitor-covid-19-update-week-1-2024. Date accessed May 30, 2025.
Continuous Mortality Investigation. (2025). Mortality Monitor. https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/other-cmi-outputs/mortality-monitor. Date accessed May 30, 2025.
GOV.UK. (2025). Excess mortality in England and English regions: March 2020 to December 2023. https://www.gov.uk/government/statistics/excess-mortality-in-england-and-english-regions. Date accessed May 30, 2025.

https://doi.org/10.1093/jrsssa/qnaf112
Advance access publication 3 July 2025

# Jonathan Rougier's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Jonathan Rougier**

School of Mathematics, University of Bristol, Bristol, UK

*Address for correspondence*: Jonathan Rougier, School of Mathematics, University of Bristol, Fry Building, Woodland Road, Bristol BS8 1UG, UK. Email: j.c.rougier@bristol.ac.uk

I would like to thank the authors for this fascinating and challenging paper. Possibly, they have been more successful than they realized, because by the time I got to their smoking gun Figure 10, which indicates that UK fatal infections peaked before the first lockdown, I was somewhat vaccinated to their conclusions.

First, Sections 4.1 and 4.2 explain how infectious epidemics have complex heterogeneous dynamics even without interventions. This made me question their assumption of a fixed time-invariant distribution for $\pi()$, the number of days from fatal infection to death, in their deconvolution approach. I appreciate that this is a common assumption, but I wondered if the authors had considered introducing an uncertain dynamic component; for example, a time-varying expectation with a constant coefficient of variation.

Second, Section 2.4 discusses the difficulty of attributing death to COVID in the presence of comorbidities. This made me wonder whether the authors could combine their nifty method for computing excess deaths adjusted by age structure with their deconvolution approach and thus infer fatal infections from reported total deaths rather than reported hospital COVID-attributed deaths. This would necessarily assume at all excess deaths in 2020 were COVID-related, but this seems to be standard, according to Section 2.4. The result would be an assessment of the dynamics of fatal infections in the country as a whole, rather than just in the hospitals.

*Conflicts of interest:* none declared.

https://doi.org/10.1093/jrsssa/qnaf126
Advance access publication 31 July 2025

# Paul Birrell, Joel Kandiah, and Daniela De Angelis' contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Paul J. Birrell[1,2], Joel Kandiah[1] and Daniela De Angelis[1,2]**

[1]MRC Biostatistics Unit, University of Cambridge, East Forvie Building, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK
[2]Analysis and Intelligence Assessment, Chief Data Officer Group, UK Health Security Agency, 61 Colindale Avenue, London NW9 5EQ, UK

*Address for correspondence*: Paul J. Birrell, MRC Biostatistics Unit, University of Cambridge, East Forvie Building, Forvie Site, Robinson Way, Cambridge CB20SR, UK. Email: paul.birrell@mrc-bsu.cam.ac.uk



Wood's deconvolution of data on deaths to estimate infection incidence is conducted at the national level. Applying it regionally offers insight into whether the national pattern masks significant spatial variation. The first wave of infection was driven by infections in London, potentially skewing the national picture. When applying Wood's deconvolution to regionally stratified data on deaths within 60 days of a positive test (Figure 1), we find that incidence in London peaked prior to the lockdown, with $\mathcal{R}_t$ falling below 1 by the time restrictions were introduced. North-East and Yorkshire shows a similar peak, but the estimated $\mathcal{R}_t$ remains above 1 at lockdown. The South-East is representative of the remaining regions: incidence declining at lockdown but with $\mathcal{R}_t > 1$ and a slower decline from a wider peak. These contradictory findings arise from the two curves (incidence and $\mathcal{R}_t$) being smoothed independently.

While real-world transmission changes around lockdown are not truly discontinuous, modelling them as such is reasonable when the rate of change in transmission is high in comparison to model time-steps. Wood's accelerated time model presented in Figure 6 of Wood (2022) suggests a steep drop in transmission just before the lockdown, a drop that eclipses any post-lockdown changes. This drop clearly represents the sum of the impacts of a succession of increasingly severe mitigation measures, including lockdown, that were steadily being imposed at this time. In the absence of high-quality prevalence data, disentangling the precise timing and true impact of these overlapping measures remains elusive.

To further explore this, we carried out a simulation study that assumed a sudden drop in transmission at the time of the lockdown, using Wood's log-normal infection-to-death distribution and a 0.0071 infection fatality rate. Simulated death data were analysed with Wood's deconvolution in the North-East and South-East. Although the underlying infection incidence peaked at lockdown, the smoothed estimates for the infection curve place the peaks 4–5 days earlier Figure 2. This demonstrates a repeatable pattern: if a lockdown were to cause such a sudden drop in transmission, Wood's method may not detect it accurately.

*Conflict of interests:* P.J.B. and D.D.A. are members of the Scientific Pandemic Infections group on Modelling ('SPI-M'), an advisory group to the UK government's Department of Health and Social Care (DHSC), designed to provide expert advice based on infectious disease modelling. The content of this response reflects personal opinions and not those of DHSC or UK Health Security Agency.

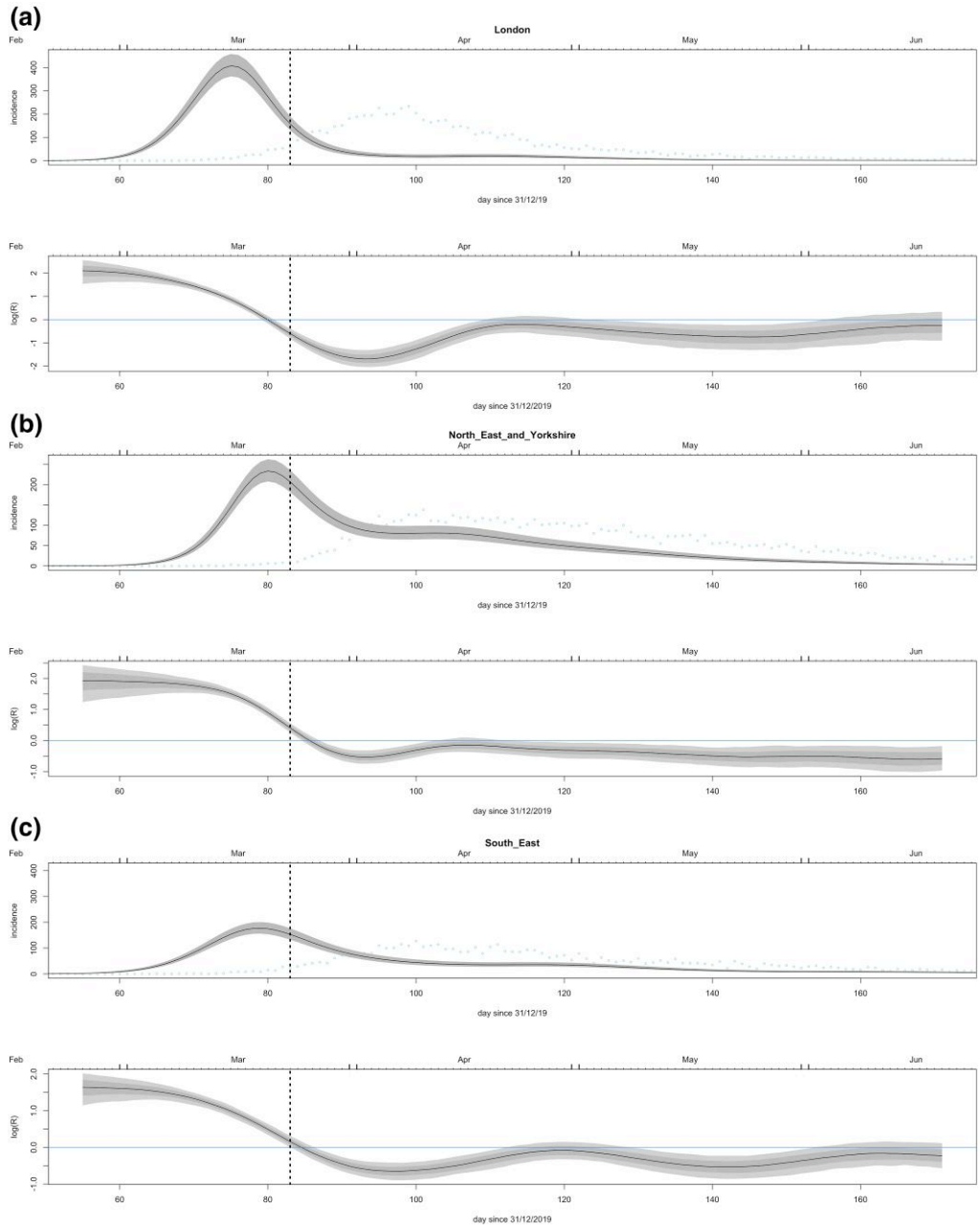



**Figure 1.** Estimated fatal incidence and $\mathcal{R}_t$ for three different regions. a) London, b) North-East and Yorkshire, and c) South-East.

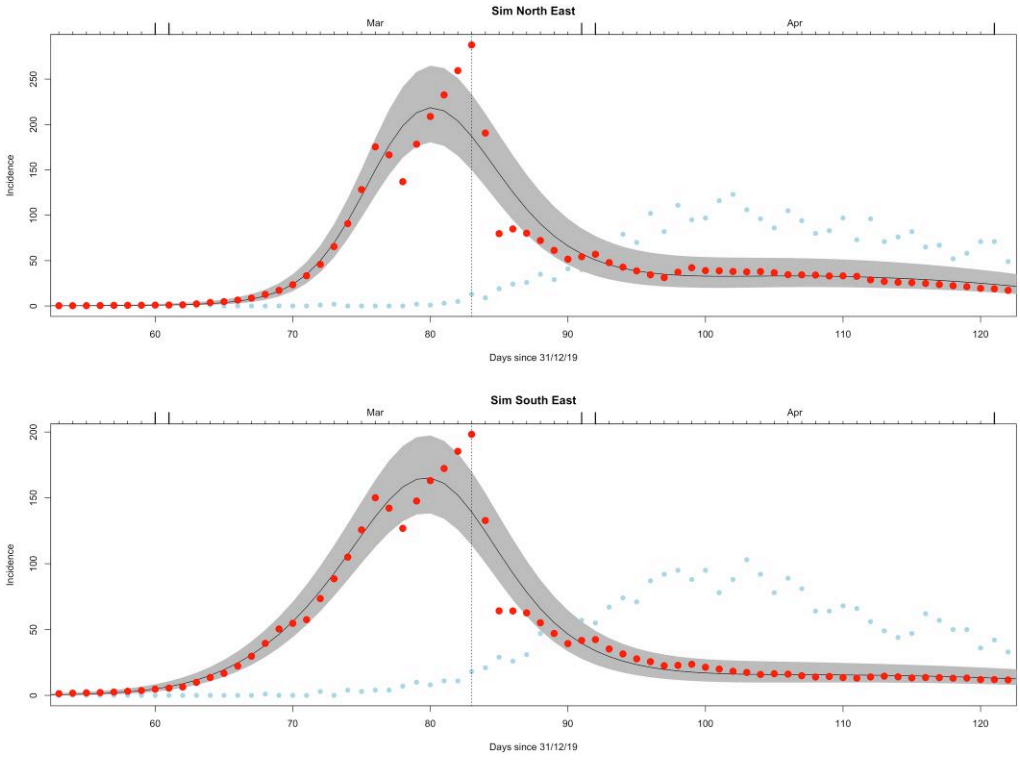



**Figure 2.** Applying the Wood deconvolution approach to estimate incidence (black line with shaded credible interval), from simulated data (blue dots). Red dots represent the underlying incidence used to generate the simulated data.

## Reference

Wood S. N. (2022, September). Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? *Biometrics*, *78*(3), 1127–1140. https://doi.org/10.1111/biom.13462

https://doi.org/10.1093/jrsssa/qnaf151
Advance access publication 10 December 2025

# Kevin Fong, Tim Cook, and Charlotte Summers' contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Kevin J. Fong[1,2,*], Tim Cook[3,4] and Charlotte Summers[5]**

[1]Department of Anaesthesia, University College London Hospitals NHS Trust, London, UK
[2]Department of Science, Technology, Engineering and Public Policy, University College London, London, UK

[3]Department of Anaesthesia and Intensive Care Medicine, Royal United Hospitals Foundation Trust, Bath, UK
[4]School of Medicine, University of Bristol, Bristol, UK
[5]Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, Cambridge, UK
*Address for correspondence*: Kevin J. Fong, Department of Anaesthesia, University College London Hospitals NHS Trust, London NW1 2BU, UK.
Email: Kevin.fong@nhs.net

## Surge capacity and the NHS during the COVID-19 pandemic

Healthcare systems have finite operating capacity. When this is substantially exceeded, the quality and safety of care is compromised, with negative consequences for patient outcomes and long-term consequences for the health service as a whole.

To illustrate the pressures faced by the NHS during COVID, we focus here on the experience of those working on intensive care units (ICUs). However, it is important to remember that most of the deaths in the NHS and social care occurred outside this setting: in the community, on general wards and in care homes, where conditions were as bad and—in many places—far worse than those seen in ICU.

In the first 18 months of the pandemic, approximately 37,000 patients, diagnosed with COVID-19, were admitted to ICUs in England. The median age of these patients was 60 years. 50% were under 60; 75% were under 69 (Intensive Care National Audit Research Centre, 2021). The scale of death in these settings was unprecedented. 40% of all patients admitted to the ICU died in the hospital within 90 days of admission (Intensive Care National Audit Research Centre, 2021).

The NHS had very little spare operating capacity immediately prior to the pandemic. In January 2020, 83% of all ICU beds were occupied with non-COVID patients. General ward bed occupancy had regularly peaked at greater than 90% in preceding years (Fong et al., 2024). To accommodate the sudden increase in demand caused by COVID-19, the NHS had to substantially modify and expand its existing capacity. The term 'ICU bed' is shorthand for a complex sociotechnical asset, which relies upon state-of-the-art technology and a large specialist team of healthcare workers. When ICUs came under pressure, it was because they were running out of specialist staff rather than furniture.

Many NHS staff were redeployed away from their ordinary duties to assist in expanding ICU and acute care. As a result, the greater the surge in COVID-19 patients, the more non-COVID care had to be shut down, leading to surgical activity reducing by nearly 10,000 cases per day during the surge in January 2021, with some specialties falling to 10% of pre-COVID activity (Kursumovic et al., 2021).

The extraordinary strain experienced by ICUs was associated with increased patient mortality and morbidity (Wilcox et al., 2022). Experiencing death at this scale, working in a cramped, less well-equipped ad hoc ICUs in which specialist staff were spread thin, took its toll.

Multiple studies show that NHS staff suffered substantial psychological harm, physical harm and—in too many cases—death during the pandemic (Andhavarapu et al., 2022; Greenberg et al., 2021; Kearney et al., 2020). During the peak of the UK's largest COVID-19 surge, in January 2021, a study of staff working on ICU showed rates of probable post-traumatic stress disorder comparable to those seen in military veterans deployed in combat roles during the war in Afghanistan (Hall et al., 2022).

The long-term effect on the NHS and its workforce has been profound. In 2022 more nurses left the NHS than at any time in its history (Palmer & Rolewicz, 2022). The NHS workforce is now measurably younger and less experienced than it was in 2015 (Rolewicz, 2024). Sickness rates are currently around 1% above the pre-COVID baseline, more than 2 years after the official end of the pandemic in the UK (NHS Digital, NHS England, 2025), alongside a substantial reduction in discretionary effort (General Medical Council, 2024; Taylor et al., 2025).

Wood et al. argue that lockdowns were 'largely unnecessary for turning around the waves of infection'. With many NHS hospitals overwhelmed at the height of the pandemic, the total patient load, peak rates of admission, peak rates of death, and peak incidence of infection all directly contributed to the severity of strain experienced by NHS services, as well as the harm to its patients

and workforce. It was important to be able to reduce patient load rapidly and definitively, to prevent NHS surge capacity from being further exceeded.

The authors agree that the extreme reduction in contact rates accompanying lockdowns suppressed transmission rates and led infections, hospital admissions and deaths to decline; which is precisely why—in the context of a service facing overwhelming pressure—this measure of last resort became necessary.

*Conflicts of interest:* During the pandemic K.J.F. served as national clinical adviser in emergency preparedness, resilience, and response for the covid-19 incident, C.S. was involved at local, regional, and national level with the NHS critical care response and along with T.C. was an author of guidelines on the covid-19 hub run by the four substantive anaesthesia and intensive care organisations. T.C. had a role in research examining both healthcare worker safety and the spread of SARS-CoV-2 infection during medical procedures. All three worked with frontline operational teams and at strategic level during the covid-19 pandemic.

## References


Andhavarapu S., Yardi I., Bzhilyanskaya V., Lurie T., Bhinder M., Patel P., Pourmand A., & Tran Q. K. (2022). Post-traumatic stress in healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. *Psychiatry Research*, *317*, 114890. https://doi.org/10.1016/j.psychres.2022.114890

Fong K. J., Summers C., & Cook T. M. (2024). NHS hospital capacity during COVID-19: Overstretched staff, space, systems, and stuff. *BMJ*, *385*, e075613. https://doi.org/10.1136/bmj-2023-075613

General Medical Council. (2024). *The state of medical education and practice in the UK. Workforce Report 2024*. General Medical Council.

Greenberg N., Weston D., Hall C., Caulfield T., Williamson V., & Fong K. (2021). Mental health of staff working in intensive care during Covid-19. *Occupational Medicine (Lond)*, *71*(2), 62–67. https://doi.org/10.1093/occmed/kqaa220

Hall C. E., Milward J., Spoiala C., Bhogal J. K., Weston D., Potts H. W. W., Caulfield T., Toolan M., Kanga K., El-Sheikha S., Fong K., & Greenberg N. (2022). The mental health of staff working on intensive care units over the COVID-19 winter surge of 2020 in England: A cross sectional survey. *British Journal of Anaesthesia*, *128*(6), 971–979. https://doi.org/10.1016/j.bja.2022.03.016

Intensive Care National Audit Research Centre. (2021). *ICNARC Report on COVID-19 in critical care: England, Wales and Northern Ireland 6 July 2021*. Intensive Care National Audit Research Centre.

Kearney L., Lennane S., Woodman E., Kursumovic E., & Cook T. M. (2020). At least 23 nationalities among NHS staff killed by covid. *The Health Service Journal*, *19*(19th May 2020).

Kursumovic E., Cook T. M., Vindrola-Padros C., Kane A. D., Armstrong R. A., Waite O., & Soar J. (2021). The impact of COVID-19 on anaesthesia and critical care services in the UK: A serial service evaluation. *Anaesthesia*, *76*(9), 1167–1175. https://doi.org/10.1111/anae.15512

NHS Digital, NHS England. (2025). *NHS Sickness absence rates*. NHS Digital, NHS England. https://digital.nhs.uk/data-and-information/publications/statistical/nhs-sickness-absence-rates/january-202. Date accessed June 10, 2025.

Palmer B., & Rolewicz L. (2022). https://www.nuffieldtrust.org.uk/resource/peak-leaving-a-spotlight-on-nurse-leaver-rates-in-the-uk2023. Date accessed May 23, 2025.

Rolewicz L. (2024). *The changing experience levels of NHS staff*. Nuffield Trust. https://www.nuffieldtrust.org.uk/resource/the-changing-experience-levels-of-nhs-staff. Date accessed June 10, 2025.

Taylor B., Scobie S., & Palmer B. (2025). *What does the NHS staff survey tell US about the changing behaviours and motivation of health care staff?* Nuffield Trust. https://www.nuffieldtrust.org.uk/news-item/what-does-the-nhs-staff-survey-tell-us-about-the-changing-behaviours-and-motivation-of-health-care-staff. Date accessed June 10, 2025.

Wilcox M. E., Rowan K. M., Harrison D. A., & Doidge J. C. (2022). Does unprecedented ICU capacity strain, as experienced during the COVID-19 pandemic, impact patient outcome? *Critical Care Medicine*, *50*(6), e548–e556. https://doi.org/10.1097/ccm.0000000000005464


https://doi.org/10.1093/jrsssa/qnaf109
Advance access publication 22 July 2025

# Sunetra Gupta's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Sunetra Gupta**

Department of Biology, University of Oxford, Oxford OX1 3SY, UK

*Address for correspondence*: Sunetra Gupta, Department of Biology, University of Oxford, Peter Medawar Building for Pathogen Research, Oxford OX1 3SY, UK. Email: sunetra.gupta@biology.ox.ac.uk



Wood et al. present a thoughtful and thorough analysis of various measurements made, and how they were interpreted, during the Covid-19 pandemic. They make a compelling case for the peak of the first wave having already been reached before lockdown measures were implemented in the majority of the 10 European countries for which they were able to obtain reliable data on daily Covid deaths. In short, the effective reproduction number had fallen below unity (or was approaching it, according to another study presented at the discussion meeting) by the time lockdowns were put in place.

This suggests that substantial resource depletion for the pathogen had occurred prior to mandated interventions, and calls into question the role that non-pharmaceutical interventions played in shaping the pandemic. It is worth considering here the various routes by which this resource depletion could have taken place in the absence of intervention.

The first, and most obvious, contribution is from the depletion of the susceptible population through acquisition of immunity—a concept known as 'herd immunity' and for which a 'herd immunity threshold' (HIT) may be defined at the point that the rate of growth of the infectious proportion of the population declines to zero (i.e. where $R = 1$). In a homogenous population this occurs roughly when the proportion immune reaches the value $1 - 1/R_0$ but, as mentioned in the article, heterogeneities in infection risk and contact patterns can have a considerable impact on HIT, as can pre-existing host resistance. It is also possible that there was a spontaneous reduction in contact rates before lockdowns were imposed. None of these complexities are, however, necessary to explain why HIT may have been reached in many settings prior to the imposition of lockdowns, as there are a variety of simple SIR scenarios involving an earlier introduction of the virus and lower infection fatality rates that are compatible with the data (https://www.medrxiv.org/content/10.1101/2020.03.24.20042291v1). Arguments against the build-up of substantial herd immunity rest on measurements of antibody positivity (as a marker of exposure) but these tend not to take into account the rapid decay of detectable antibodies (https://pubmed.ncbi.nlm.nih.gov/39827808/); neither do they acknowledge that not all individuals mounted an antibody response upon first exposure https://www.thelancet.com/action/showPdf?pii=S2352-3964%2824%2900511-5). Finally, it should be recognized that seasonality played a large role in the dynamics of Covid-19 (https://www.science.org/doi/epdf/10.1126/science.abd7343) and that most epidemiological patterns worldwide can be fit at the appropriate scale to an SIRS model with seasonality.

*Conflicts of interest:* none declared.

**The following contributions were received in writing after the meeting:**

https://doi.org/10.1093/jrsssa/qnaf101
Advance access publication 7 July 2025

# Adrien Allorant's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Adrien Allorant**

Department of Social Statistics and Demography, University of Southampton, Southampton, UK

*Address for correspondence:* Adrien Allorant, Department of Social Statistics and Demography, University of Southampton, University Rd, Highfield Campus, Southampton SO17 1BJ, UK. Email: adrien.allorant@soton.ac.uk



Wood et al. offered a valuable perspective on COVID-19 response strategies, including a preliminary cost–benefit assessment. Nonetheless, several points warrant further scrutiny. First, their estimate of life-years saved accounts only for mortality while excluding morbidity—such as long COVID and post-infection cognitive impairment—which can reduce quality of life and exert substantial economic and healthcare burdens. Moreover, the mortality figures themselves may be conservative, as indicated by the WHO Excess Mortality Team's Nature study suggesting that true COVID-19 death tolls outstrip official counts (Msemburi et al. 2023). Such underestimation risks skew the cost-per-life-year analysis upward.

Second, while Figure 1 clarifies that COVID-19 contributed a small share of total deaths in younger populations, it does not imply a low case fatality rate for those groups. Historically, all-cause mortality among the young is low, so even a modest infection fatality rate can produce significant excess deaths when large cohorts are exposed.

Third, the paper's argument that economic shocks may ultimately cause more life loss than COVID itself rests on historical recession data that may not extrapolate cleanly to the UK pandemic context. Crises in countries without universal health coverage have shown strong links between economic hardship and mortality. Yet in the UK, extensive fiscal measures, furlough schemes, and the NHS likely mitigated the worst outcomes. Consequently, interpreting the 2008 global financial crisis as a straightforward analogue could overstate the indirect health costs of COVID-19 interventions.

Finally, Wood et al. provided insufficient attention to the imminent threat of critical care overload. Nonpharmaceutical interventions aimed to reduce ICU saturation, an acute danger during infection surges. Had the NHS been overwhelmed, mortality from both COVID-19 and other conditions would have escalated sharply (Fong et al. 2024). Assessing the cost-effectiveness of lockdowns and related policies should factor in the potential collapse of healthcare delivery, which carries profound human and economic ramifications.

In summary, while Wood et al. stimulate a valuable debate, future analyses should incorporate morbidity measures, robust excess mortality estimates, careful evaluation of recession analogies, and the imperative of safeguarding healthcare capacity. Such a holistic approach would yield more balanced conclusions about the trade-offs entailed in pandemic policymaking.

*Conflicts of interest:* none declared.

## Data availability

No new data were generated or analysed in support of this research.

## References

Fong, K. J., Summers, C., & Cook, T. M. (2024). NHS hospital capacity during COVID-19: Overstretched staff, space, systems, and stuff. *BMJ*, *385*, e075613. https://doi.org/10.1136/bmj-2023-075613

Msemburi, W., Karlinsky, A., Knutson, V., Aleshin-Guendel, S., Chatterji, S., & Wakefield, J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, *613*, 130–137. https://doi.org/10.1038/s41586-022-05522-2

https://doi.org/10.1093/jrsssa/qnaf099
Advance access publication 7 July 2025



# Luke A. Barratt's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Luke A. Barratt**

Statistical Laboratory, University of Cambridge, Centre for Mathematical Statistics, Wilberforce Road, Cambridge, CB3 0WB, UK

*Address for correspondence*: Luke A. Barratt, Statistical Laboratory, University of Cambridge, Centre for Mathematical Statistics, Wilberforce Road, Cambridge, CB3 0WB, UK. Email: lab85@cam.ac.uk

It is well noted in the paper that it is crucial to understand and include heterogeneities in the underlying dynamics to fully appreciate the progression of an epidemic. In particular noted is the fact that individuals do not have uniform social interaction and therefore are not uniformly exposed nor uniformly spread the disease. However, little is made in the statistical analysis of geographical heterogeneity. In particular, in the case of the Covid-19 pandemic, it was noted at the time that London was extraordinary in its experience of the first wave in 2020. Indeed, this fact has been demonstrated statistically by many authors, including recently by Barratt and Aston in two preprints (Barratt & Aston, 2024, 2025) which make minimal modelling assumptions. It is not clear that the analysis provided on the aggregate level for the UK is valid for the various constituent nations and regions without explicit consideration of local effects.

This raises three main questions:

1. Does the analysis when disaggregated by region demonstrate that R was indeed below 1 by the time of the lockdown in all regions? Barratt and Aston (2024) have demonstrated that London saw a particularly early peak, and this could be overwhelming the analysis. Moreover, it is not clear that the dynamics in London could be extrapolated to other regions. Given there are substantial local impacts of incidence on, for example, hospital capacity, it is crucial to consider the dynamics at a more local scale.
2. On the analysis of non-pharmaceutical interventions, namely lockdown, it appears that the main focus has been on the first national lockdown. It would be beneficial to hear from the authors whether consideration of targeted interventions in local areas, as became standard later on in the pandemic, can be analysed in the same way, and whether the same conclusions would be drawn. Once again, the analysis of the latter restrictions seem to have been based on an aggregate view of the dynamics.
3. When considering the communication of statistics, various dashboards were available mapping local levels of Covid-19 incidence. Anecdotally, this seemed to be followed quite closely by the general public, perhaps aiding in both the public's perception of local risk and locally influencing behaviour in response to said risk. What considerations do the authors have around this local communication of risk?

*Conflicts of interest:* No conflicts of interest to declare.

## References

Barratt L. A., & Aston J. A. D. (2025, December). 'Exploring spatiotemporal variation in Covid-19 waves: Non-Euclidean spatially aware functional registration.' *Ann. Appl. Stat. 19*(4), 3261–3281. https://doi.org/10.1214/25-AOAS2077

Barratt L. A., & Aston J. A. (2025). 'A nonparametric and functional wombling methodology' arXiv, arXiv:2502.07740, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2502.07740

https://doi.org/10.1093/jrsssa/qnaf118
Advance access publication 11 July 2025



# Tom Britton's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Tom Britton**

Department of Mathematics, Stockholm University, Stockholm, Sweden

*Address for correspondence:* Tom Britton, Department of Mathematics, Stockholm University, SE-10691 Stockholm, Sweden. Email: tom.britton@math.su.se

I thank the authors for contributing to the aftermath of how data were analysed and presented for the general public during the pandemic, with particular focus on UK. Below I list some points that I find interesting in the paper, adding some additional viewpoints.

## Risks

The authors make a correct point that the individual risk for younger people was, at best, not explicitly stated to be very low, or at worst, intentionally hidden. And, rather than using 'hard hitting emotional messaging' a better strategy should be the honest path to state that young people are not at high risk of becoming seriously ill, but that their preventive efforts are highly important for reducing spreading and thus protecting risk groups and the older population.

## Competing risks of life loss

The authors bring up the important point to compare costs/risks for the severe infections with those of the effect of preventive measures. My personal opinion is that such discussions have received far too little attention, also in the aftermath. It is of course very hard to estimate the cost of severe infections and the cost to society in general, under different preventive strategies, but even if hard this has to be done during a pandemic. The cost to society from preventive measures received way less attention compared with the cost of severe infections. The authors estimate that the cost, in terms of loss of life-years which seems like the reasonable cost to compare, was 3–4 higher from the restrictions compared with the life-years loss from COVID. The computations seem reasonable, but it would be interesting to see estimates from other groups of scientists.

## Cost-effectiveness of different preventive measures

A related important research questions which should be addressed more is the cost-effectiveness of different preventive measures. This is of course hard, but even rough estimates of which measures give the most bang for the buck would be very welcome. If such a list is agreed upon it would give

an order at which preventive measures are sequentially inserted up until the overall cost exceeds that of the cost of severe infections from an epidemic. For instance, I would guess that vaccination (once they exist) and face-masks give most bang for the buck and perhaps school closure being at the other end of the scale.

## Individual heterogeneities

The authors correctly point out that individual heterogeneities were not acknowledged enough in most models used for predictions. Such heterogeneities are substantial in reality as, for example, manifested in social contact studies (Britton & Ball, 2025), and it is well known from numerous models that heterogeneities imply substantially smaller outbreaks than when predicting using homogeneous models. Another effect of neglecting heterogeneities is that reduction in spreading may be attributed to successful preventive measures, whereas most of the reduction may be explained by socially active by then being infected and immune.

## Involvement from academia

As an infectious disease modeller, there is another, more positive impression that I have of the UK COVID-19 strategy not mentioned in the paper, and that is the strong involvement of the scientific community (SAGE, Spi-M, reference groups, etc.). In Sweden, and many other countries, scientists were more or less left out in the process of discussing strategies and making predictions. I think the (better) UK situation was mainly due because many of these networks were in place already prior to arrival of the pandemic. Many other countries, including my own, still have the homework to prepare such collaborative groups before the next crisis—but in UK the relevant infrastructure seem already to be in place.

*Conflicts of interest:* none declared.

## Reference

Britton, T., & Ball, F. (2025). Improving the use of social contact studies in epidemic modelling. *Epidemiology*. https://doi.org/10.1097/EDE.0000000000001876

https://doi.org/10.1093/jrsssa/qnaf069
Advance access publication 27 June 2025

# Fergus Chadwick, Dirk Husmeier, and Jason Matthiopoulos's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Fergus Chadwick[1], Dirk Husmeier[2] and Jason Matthiopoulos[3]**

[1]School of Mathematics & Statistics, University of St Andrews, St Andrews, UK
[2]School of Mathematics and Statistics, University of Glasgow, Glasgow, UK
[3]School of Biodiversity, One Health & Veterinary Medicine, University of Glasgow, Glasgow, UK

*Address for correspondence*: Dirk Husmeier, The Mathematics and Statistics Building, University of Glasgow, University Place, Glasgow G12 8SQ, UK. Email: dirk.husmeier@glasgow.ac.uk

We congratulate the authors on a thought-provoking article that stimulates constructive methodological discourse and reflection on our work to support decision-makers during the pandemic. We

would like to contribute some thoughts on the topic of life loss estimates. The authors have criticized the treatment of comorbidities as simple categories in the study by Hanlon et al. (2020) and claim that different life loss estimates may have been obtained if comorbidity severity indicators had been included. The uncertainty of the life loss estimates would certainly have been reduced by such an inclusion, due to the well-known fact that upon conditioning on additional factors, the conditional Shannon entropy of a response (here: years of life lost) is reduced. The authors argue that a sub-categorization of the comorbidity classes used in Hanlon et al. (2020), according to the degree of severity, would most likely have found an over-representation of Covid deaths in the sub-categories corresponding to higher severity levels. This, in turn, implies that a lower estimate of the expected years of life lost would have been obtained (assuming severity always means a shorter life expectancy). However, data for such sub-categories were not available at that time, and the fact that the authors refer to them in hypothetical terms rather than with specific reference to a data repository suggests that they are not currently available either. As the study by Hanlon et al. (2020) included several cases of multiple comorbidities, it could be argued that the severity of each individual comorbidity was at a lower degree of the severity spectrum. Otherwise, a patient may have succumbed to their combined effect prior to contracting Covid, or indeed not have been as likely to be exposed to Covid in the early stages of the pandemic. This would lead to a conclusion opposite to that arrived at by the authors; however, like the authors' statement, this is a conjecture that is not supported by any data. We would argue that any decision-making process in times of a global crisis should only be based on data and facts available at the time a decision has to be made—not hypothetical scenarios that can only be verified in the future. The latter should obviously be included as soon as they can be substantiated with actual data, which may then lead to a revised risk assessment and different scenario planning.

*Conflicts of interest:* none declared.

## Reference

Hanlon, P., Chadwick, F., Shah, A., Wood, R., Minton, J., McCartney, G., Fischbacher, C., Mair, F. S., Husmeier, D., Matthiopoulos, J., & McAllister, D. A. (2021). COVID-19—Exploring the implications of long-term condition type and extent of multimorbidity on years of life lost: A modelling study. *Wellcome Open Research*, *5*, 75. https://doi.org/10.12688/wellcomeopenres.15849.3

https://doi.org/10.1093/jrsssa/qnaf116
Advance access publication 23 July 2025

# Dr Arun Peter Chind's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Arun Chind**

Statistician, Econometrian & Director, Proshen Health & Risk Consulting, Wakefield, UK

*Address for correspondence:* Arun Chind, Statistics and Econometrics Section, Proshen Consulting, 87 Dacre Avenue, Wakefield WF2 8AH, UK. Email: arunpeter@yahoo.com

While it is appealing to discuss ways of doing better 'next time', it is helpful to remind ourselves that the Covid-19 pandemic was both an extreme and infrequent event and therefore, defied prediction or perfect planning. Consider the annual hurricane season when despite forewarning cloud patterns nonetheless leads to loss of life and damage to property. Discussion around excessive spending and the long-term effects on the economy can be seen as an exemplar of the *winner's*

*curse* in situations of information asymmetry recognized by economists. Also termed *winner's remorse*, this term is used to denote the phenomenon where the winning bidder at an auction experiences remorse upon reflection, realizing that she has paid too much. In this light, one must surely ask—are we throwing good money (recall that today, is a 'present') after bad (the sunk cost—money already spent, and the time being wasted in self-recrimination).

One of the points raised in this article is the NICE threshold for approval of an intervention (about £30,000 per QALY). No doubt, this threshold can be helpful in deciding whether or not to fund an intervention for a single patient (e.g. cancer), given competing demands on the money in other areas (e.g. for cystic fibrosis). However, the use of such an arbitrary threshold for the 'cost–benefit analysis' of an extreme and infrequent event of unknown course and consequence, in a public health crisis with its ramifications including network effects on other parts of the economy, is very questionable.

One of the matters raised in the discussion related to care workers going into people's homes/working at care facilities to provide personal care (feeding, washing etc.) and ending up spreading Covid-19, and likely leading to the deaths of vulnerable residents. Care workers are often not paid a lot more than the minimum wage to perform duties that most of us would not bear to ponder upon. Those in stable employment exploited the 'shielding advice' arising from a diagnostic label or connection to a person with a diagnostic label, to be paid for staying home. This left gaps that were filled with temporary workers. Such workers are paid more on 'temporary' contracts that involve itinerant work, which incentivizes them to work in this shady sector (liable to government-sanctioned immoral/unethical practices) of the economy. The unintended victim-shaming of care workers engaged in precarious work, I found the most uncomfortable part of the meeting.

*Conflicts of interest:* Dr Chind is a paid employee and director of Proshen Consulting, a consultancy that offers research services and policy advice based on sound principles of economics and law, to client organisations. Using his training in statistics (GradStat/GStat), economics (including qualifications from UEA and Purdue) and training in law (including an LLM from Leeds Beckett), he advocates for the improvement of working conditions for workers. Dr Chind has the lived experience of working as a 'Locum' Consultant Occupational Physician at several NHS Trusts, under precarious work conditions.

https://doi.org/10.1093/jrsssa/qnaf090
Advance access publication 28 June 2025

# John Dagpunar's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al

**John Dagpunar**

School of Mathematics, University of Southampton, Southampton, UK

*Address for correspondence*: School of Mathematics, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK. Email: j.dagpunar@soton.ac.uk

I congratulate the authors on their wide-ranging paper and the careful analysis therein. Their conclusion that fatal infection incidence $f_t$ had peaked before compulsory lockdowns started in the majority of European countries is significant. However, that does not necessarily mean that lockdowns should not have happened before reaching a peak if infective prevalence had reached dangerous levels. Equally, one needs to model whether voluntary or behavioural suppression by a population is sufficient to justify no compulsory lockdown.

Taking the approach used for estimating the effective reproduction number using inferred infection incidence estimates $\widehat{f}_t$, let us assume, but only for the period up until vaccination became available, a heterogeneous SEIR model $\frac{dS}{dt} = -\delta R_e(t)I(t) = -\delta S(t)^{\lambda} R_c(t)I(t)$, $\frac{dE}{dt} = \delta R_e(t)I(t) - \gamma E(t)$, $\frac{dI}{dt} = \gamma E(t) - \delta I(t)$, where, $\gamma^{-1} = 3$ and $\delta^{-1} = 5$ days as in Wood (2021), $\lambda$ takes a value in the range referred to in the paper, and where $R_e(t)$ and $R_c(t)$ denote the effective and control reproduction numbers, respectively. We have $R_e(t) = S(t)^{\lambda} R_c(t)$. Replacing $\delta R_e(t)I(t)$ with $\widehat{f}_t$, the solution to the last two differential equations is for $\delta \neq \gamma$, $E(t) = \int_0^t e^{-\gamma(t-u)} \widehat{f}_u du$, $I(t) = \frac{\gamma}{\delta-\gamma} \int_0^t \left(e^{-\gamma(t-u)} - e^{-\delta(t-u)}\right) \widehat{f}_u du$. From the first differential equation $\widehat{f}_t = \delta R_e(t)I(t)$ or $R_e(t) = \frac{\widehat{f}_t}{\delta I(t)}$.

This enables one to examine policies alternative to the historical ones. As a first example, suppose a country's $n^{th}$ official/compulsory lockdown occurred in $(t_{n-1}, t_n]$. What might have been the trajectory of infections during that period, conditional upon taking the previous history of infections/lockdowns as a given, if the lockdown had instead been replaced by NPIs of a strength present immediately before $t_{n-1}$? One could examine such a policy by setting the initial condition in the SEIR model as $S(t_{n-1}) = 1 - \int_0^{t_{n-1}} \widehat{f}_u du$, $I(t_{n-1}) = \frac{\gamma}{\delta-\gamma} \int_0^{t_{n-1}} \left(e^{-\gamma(t_{n-1}-u)} - e^{-\delta(t_{n-1}-u)}\right) \widehat{f}_u du$, $E(t_{n-1}) = \int_0^{t_{n-1}} e^{-\gamma(t_{n-1}-u)} \widehat{f}_u du$, together with controls $R_c(t) = \frac{R_e(t_{n-1})}{S(t_{n-1})^{\lambda}} = \frac{\hat{f}_{t_{n-1}}}{\delta I(t_{n-1}) S(t_{n-1})^{\lambda}}$ for $t_{n-1} < t \leq t_n$.

As a second example, given that infection incidence peaked before a first lockdown, would it in retrospect have been better to implement the lockdown earlier and arguably before reaching that peak? Suppose a compulsory lockdown of duration $t_1 - t_0$ started not at the historical $t_0$ but at $t'_0$, where $t'_0 < t_0$. Then, one could model the behaviour where lockdown of the same duration and strength as the historical one, now takes place during $[t'_0,\ t'_0 + t_1 - t_0)$ followed by the interval $[t'_0 + t_1 - t_0,\ t_1]$ where the controls replicate the historical ones preceding the historical lockdown. We would set the initial condition to $S(t'_0) = 1 - \int_0^{t'_0} \widehat{f}_u du$, $I(t'_0) = \frac{\gamma}{\delta-\gamma} \int_0^{t'_0} \left(e^{-\gamma(t'_0-u)} - e^{-\delta(t'_0-u)}\right) \widehat{f}_u du$ , $E(t'_0) = \int_0^{t'_0} e^{-\gamma(t'_0-u)} \widehat{f}_u du$, together with controls

$$R_c(t) = \frac{\hat{f}_{t+t_0-t'_0}}{\frac{\gamma\delta}{\delta-\gamma} \int_0^{t+t_0-t'_0} \left(e^{-\gamma(t+t_0-t'_0-u)} - e^{-\delta(t+t_0-t'_0-u)}\right) \widehat{f}_u du \left(1 - \int_0^{t+t_0-t'_0} \widehat{f}_u du\right)^{\lambda}} \text{ when } t'_0 \leq t < t'_0 + t_1 - t_0 \text{ and}$$

$$R_c(t) = \frac{\hat{f}_{t-t_1+t_0}}{\frac{\gamma\delta}{\delta-\gamma} \int_0^{t-t_1+t_0} \left(e^{-\gamma(t-t_1+t_0-u)} - e^{-\delta(t-t_1+t_0-u)}\right) \widehat{f}_u du \left(1 - \int_0^{t-t_1+t_0} \widehat{f}_u du\right)^{\lambda}} \text{ when } t'_0 + t_1 - t_0 \leq\ t \leq t_1.$$

In conclusion, the authors' results for several European countries in 2020 can be used to see what the effect of different suppression policies might have been, compared with the historical record. In turn, this can inform the debate about balancing surging infection levels against the now well-known economic, societal, and health damage that follows from compulsory lockdowns.

*Conflicts of interest:* none declared.

https://doi.org/10.1093/jrsssa/qnaf087
Advance access publication 8 July 2025



# Giacomo De Nicola's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Giacomo De Nicola**

Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA, USA

*Address for correspondence*: Giacomo De Nicola, Department of Nutrition, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115, USA. Email: gdenicola@hsph.harvard.edu

I commend the authors for their sharp, thorough, and forthright examination of statistical issues surrounding the UK's Covid-19 response. Their critique spans both methodological and communicative failures, and I agree with several of their key points, including the distortion of personal risk perception, the disproportionate emphasis on worst-case modelling, and the failure to incorporate the broader risks and costs of non-pharmaceutical interventions into formal decision-making. As the authors rightly emphasize, these consequences—ranging from delayed healthcare to economic dislocation, itself contributing to life loss—should have been weighed alongside the direct risks posed by the virus.

Although the paper focuses on the UK, the practices it critiques have been widespread across high-income countries. The fact that these approaches were near-universal suggests that the problems were not limited to a few specific governments but reflected broader institutional blind spots. This makes it all the more important to draw careful lessons.

The paper touches on many important issues—far more than I can address here. I choose to focus on the estimation of excess mortality, a central concern throughout the pandemic, and one where I strongly support the authors' argument that proper age adjustment is essential. Like them, I am struck by how widely demographic fundamentals continue to be overlooked in many national and global excess deaths estimates. In work focused on Germany (De Nicola et al., 2022), we introduced a life-table-based method for calculating expected mortality that explicitly accounts for changes in population age structure, showing how failure to do so can substantially inflate excess death estimates. We later extended this approach in a follow-up study—published in this very Journal—incorporating a non-parametric uncertainty framework based on historical mortality variation, and applying the method to 30 high-income countries (De Nicola & Kauermann, 2025). Across both studies, we highlight how commonly used and seemingly 'model free' approaches—typically based on multi-year averages—tend to overstate excess mortality in ageing or demographically irregular populations. The core insight, shared with the authors, is that demographic change is not a nuisance to be averaged out, but a central element in any meaningful mortality analysis.

Balancing urgency with analytical care is never easy during a crisis. Acknowledging the human cost of the pandemic is essential, but so too is recognizing the complex and often long-term consequences of the response itself. In that context, striving for statistical honesty—even when findings are uncomfortable or politically inconvenient—is crucial for maintaining public trust. The authors' work sets a valuable benchmark for how pandemic impact and response should be assessed—one that I hope will inform both research and policymaking in future crises.

*Conflicts of interest:* none declared.

## References

De Nicola, G., & Kauermann, G. (2025). Estimating excess mortality in high-income countries during the COVID-19 pandemic. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *188*(1), 205–222. https://doi.org/10.1093/jrsssa/qnae031

De Nicola, G., Kauermann, G., & Höhle, M. (2022). On assessing excess mortality in Germany during the COVID-19 pandemic. *AStA Wirtschafts-und Sozialstatistisches Archiv*, *16*(1), 5–20. https://doi.org/10.1007/s11943-021-00297-w

https://doi.org/10.1093/jrsssa/qnaf102
Advance access publication 1 July 2025

# Peter J Diggle's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Peter J. Diggle***

Medical School, Lancaster University, Lancaster, UK

*Address for correspondence*: Peter J. Diggle, Medical School, Lancaster University, Lancaster, LA1 4YB, UK.
Email: p.diggle@lancaster.ac.uk



Today's paper is an invaluable contribution at what should be a time not only to reflect but also to prepare for the next pandemic.

One statistical aspect of the COVID-19 epidemic in the UK that the authors do not discuss, and on which much of the modelling literature that they review is also silent, is spatial heterogeneity. As early as May 2020, the REACT study led by Prof Paul Elliott at Imperial College showed more than a threefold variation in prevalence at regional level, from 0.06% in the South West to 0.2% in London. Spatial heterogeneity persisted, not only at regional level, but on much smaller spatial scales; see, for example, Figures 3–5 of Fry et al. (2021) or Figures 5 and 6 of Nicholson et al. (2022). When coupled with the use of nonlinear models, spatial heterogeneity casts doubt on the validity of findings, notably with regard to the much-quoted R-number, that are based only on national-level data. Teh et al. (2022) described a method for estimating a local (in space and time) R-number by treating $R(x,t)$ as the realization of an unobserved spatio-temporal stochastic process. Nicholson et al. (2022) estimated $R(x,t)$ by fitting SIR models to de-biased LTLA-level data. The Society held an extended discussion meeting over 2 days in June 2021 to consider this and related issues; for an introduction and references, see Diggle and Richardson (2022).

A minimal aspiration for a wealthy country in the twenty-first century should be a public health surveillance system that can deliver predictive inferences on key health indicators at fine spatial resolution in near-real time. The necessary statistical and technological tools have been available for more than 20 years (Diggle et al., 2003).

*Conflicts of interest:* No conflicts of interest.

## References

Diggle, P., Knorr-Held, L., Rowling, B., Su, T., Hawtin, P., & Bryant, T. (2003). On-line monitoring of public health surveillance data. In R. Brookmeyer, & D. F. Stroup (Eds.), *Monitoring the health of populations: Statistical principles and methods for public health surveillance* (pp. 233–266). Oxford University Press.

Diggle, P. J., & Richardson, S. (2022). Introduction. *Journal of the Royal Statistical Society A*, *185*(Suppl 1), S3–S4. https://doi.org/10.1111/rssa.12883

Fry, R. J., Hollinghurst, J., Stagg, H. R., Thompson, D. A., Fronterre, C., Orton, C., Lyons, R. A., Ford, D. V., Sheikh, A., & Diggle, P. J. (2021). Real-time spatial health surveillance: Mapping the UK COVID-19 epidemic. *International Journal of Medical Informatics*, *149*, Article 104400. https://doi.org/10.1016/j.ijmedinf.2021.104400

Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K. B., Jersakova, R., Lomax, J., King, R. E., Mallon, A.-M., Diggle, P. J., Richardson, S., Blangiardo, M., & Holmes, C. (2022). Local prevalence of transmissible SARS-CoV-2 infection: An integrative causal model for debiasing fine-scale targeted testing data. *Nature Microbiology*, *7*(1), 97–107. https://doi.org/10.1038/s41564-021-01029-0

Teh, Y. W., Elesedy, B., He, B., Hutchinson, M., Zaidi, S., Bhoopchand, A., Paque, U., Tomasev, N., Read, J., & Diggle, P. J. (2022). Efficient Bayesian inference of instantaneous reproduction numbers at fine spatial scales, with an application to mapping and nowcasting the COVID-19 epidemic in British local authorities. *Journal of the Royal Statistical Society A*, *185*(Suppl 1), S65–S85. https://doi.org/10.1111/rssa.12971

https://doi.org/10.1093/jrsssa/qnaf078
Advance access publication 30 June 2025

# Erengul Dodd, Jonathan J. Forster, and Jason Hilton's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Erengul Dodd[1], Jonathan J. Forster[2] and Jason Hilton[1]**

[1]Centre for Population Change, University of Southampton, Southampton, UK
[2]Department of Statistics, University of Warwick, Coventry, UK

*Address for correspondence*: Jonathan J. Forster, Department of Statistics, University of Warwick, Coventry CV4 7AL, UK. Email: J.J.Forster@warwick.ac.uk



In Section 2.4, the authors consider the computation of excess deaths, defined as *the excess of actual deaths over what might reasonably be expected given the situation in previous years*. Equivalently, we might characterize this as excess of actual deaths over what might reasonably have been forecast given mortality experience up to the start of the pandemic. We argue that it is natural, therefore, to use mortality forecasting methodology for this computation. This approach has the advantage both of allowing for predictable variation in exposures to risk at different ages over time, as the authors do, but also of incorporating predictable patterns of mortality change, which the authors do not. Mortality forecasting in the UK context generally assumes that mortality will improve, in line with experience over recent decades and this mortality improvement is embedded into forecasting methodology. Different models for weekly Covid excess mortality have been proposed by, for example, Aburto et al (2021), the EuroMOMO group (e.g. Nielsen et al, 2021), Ellison et al (2021) and OHID (2022), but all incorporate some combination of seasonal effects, holiday effects and, critically, a log-linear mortality improvement trend based on historical data. One key difference is over what period that historical trend is estimated. Our approach (Ellison et al., 2021) is to assume that the long-term (60-year) trend applies, but the model also incorporates a sporadic, unpredictable (age-independent) mortality improvement component, in common with the classic mortality forecasting models of Lee and Carter (1992) and Cairns et al. (2011). This latter component contributes considerably to the uncertainty associated with any forecast and hence any estimate of Covid excess mortality. Our estimates, an updated version of the original estimates in Ellison et al. (2021), are displayed in Figure 1. The estimated excess mortality is somewhat larger than the authors' estimates, due to the allowance for mortality improvement in the forecast model but are similar to the ONS estimates. The ONS approach also ignores mortality trends, and additionally fails to account for ageing population, as pointed out by the authors. It seems that these two limitations, working in opposing directions, almost exactly cancel each other out, the two approaches in Figure 1 yielding similar results despite their somewhat differing assumptions.

The significant uncertainty resulting from the sporadic component in the forecasting model is noteworthy and suggests that precise inference about this aspect of the pandemic may be elusive.

## Data availability

The data underlying this article are available from the Human Mortality Database (https://www.mortality.org) and the Office for National Statistics (https://www.ons.gov.uk).

*Conflicts of interest:* None declared.

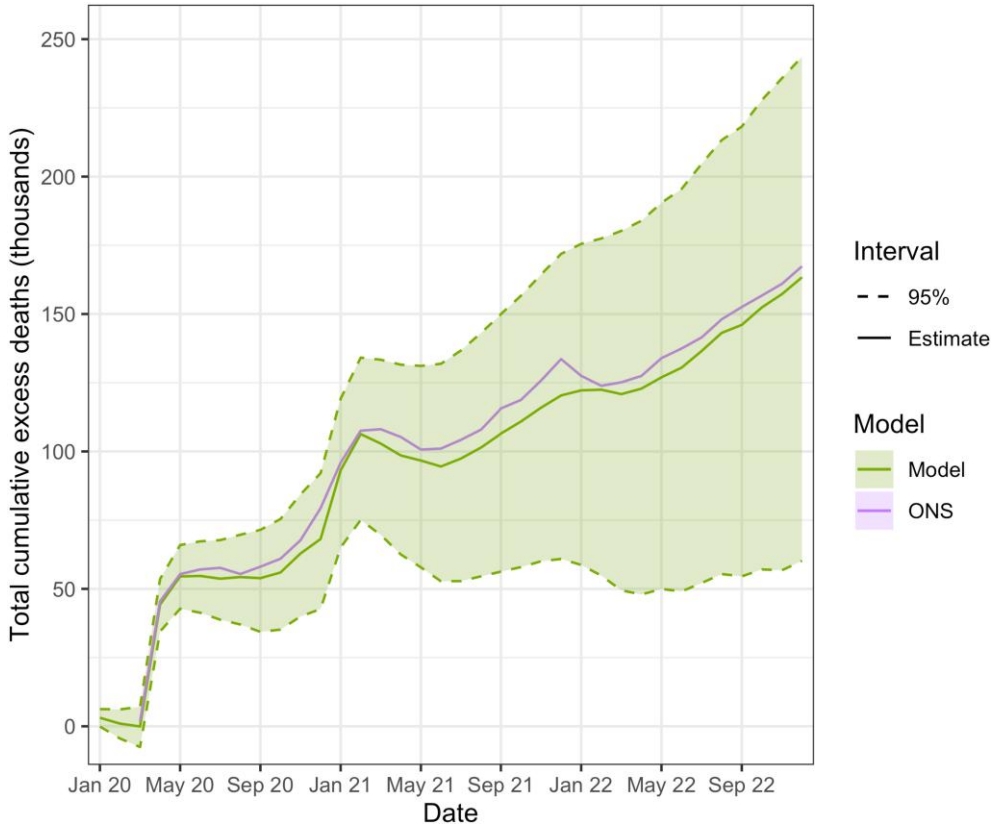



**Figure 1.** COVID excess mortality estimates and uncertainty, compared with ONS.

## References

Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., Mills, M. C., & Dowd J. B., (2021). Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: A population-level analysis. *Journal of Epidemiology and Community Health*, *75*(8), 735–740. https://doi.org/10.1136/jech-2020-215505

Cairns, A. J. G., Blake, D., Dowd, K., Coughlan, G. D., Epstein, D., & Khalaf-Allah, M. (2011). Mortality density forecasts: An analysis of six stochastic mortality models. *Insurance: Mathematics and Economics*, *48*, 355–367. https://doi.org/10.1016/j.insmatheco.2010.12.005

Ellison, J., Hilton, J., Dodd, E., Forster, J. J., Smith, P. W. F., & Bijak, J. (2021). Modelling weekly mortality incorporating annual trends. Presented to International Population Conference 2021. https://ipc2021.popconf.org/uploads/210936

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, *87*, 659–675. https://doi.org/10.2307/2290201

Nielsen, J., Nørgaard, S. K., Lanzieri, G., Vestergaard, L. S., & Moelbak, K. (2021). Sex-differences in COVID-19 associated excess mortality is not exceptional for the COVID-19 pandemic. *Scientific Reports*, *11*(1), 20815. https://doi.org/10.1038/s41598-021-00213-w

OHID. (2022). *Excess mortality in England: methodology*. https://fingertips.phe.org.uk/documents/EMMethodology.pdf Accessed 9 May, 2025.

https://doi.org/10.1093/jrsssa/qnaf103
Advance access publication 7 July 2025

# Louise Dyson's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Louise Dyson**

Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, Mathematics Institute and School of Life Sciences, University of Warwick, Coventry, UK

*Address for correspondence*: Louise Dyson, Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL, UK. Email: l.dyson@warwick.ac.uk



I will concentrate here on two main points, and also note the importance of what data was available contemporaneously and how quickly.

**Section 1.** It is extremely difficult, even in retrospect, to robustly estimate the impact of COVID-19 control measures: firstly, because multiple measures were implemented simultaneously; secondly, because the public reaction to those measures is uncertain; and thirdly, because it requires identifying the counterfactual scenario that would have happened without these measures. Therefore, it is naïve to assume that the number of 'lives saved' by lockdown can be calculated by using the Ferguson et al. prediction for a no-mitigation scenario (Ferguson et al., 2020) (for a single variant wave) minus the number of recorded deaths.

Determining counterfactual predictions is complicated by the public's reaction to increasing deaths, the effect on the healthcare system (impacting death rates due to COVID-19 and other causes), the emergence of multiple new variants and the rollout of vaccines. This is likely to remain a major topic of active research over the coming years. In addition, Wood et al.'s assertion (in the discussion) that a 'red-team' would have questioned the cost being 'many times the NICE threshold', is debatable, as the calculation made in the introduction (even if we accept it as valid) could not have been made in March 2020. Indeed, one might argue that HM Treasury acted as 'red-team' for much of the pandemic, providing an economic counterpoint to public health considerations. It is a shame that the Treasury's perspectives remain largely opaque to the general public.

**Section 2.2.** While the authors' critique is that the economic consequences of control were never factored into calculations of health risk, conversely the authors have not considered the consequences of a large and uncontrolled outbreak on the economy. In our own work (Tildesley et al., 2022), we find that while the March 2020 lockdown had a massive impact on the country's gross domestic product (GDP), the circuit-breaker lockdown of November 2020 and the lockdown of early 2021 had far less impact on the economy. This begs the question: how much were the economic shocks due to UK controls and how much was due to widespread global uncertainty? Similarly, although contact rates were reducing prior to government rules, it is unclear what would have happened without a 'lockdown' order. Removing the government rules and advice would also have precluded government financial support, leading to (even more) inequality in who could reduce their (and other's) risk.

*Conflict of interest:* I was a member of the modelling subgroup of UK SAGE (SPI-M-O) during the COVID-19 pandemic, and am currently a member of SPI-M.

## References

Ferguson, N. M., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., Dighe, A., Dorigatti, I., Fu, H., Gaythorpe, K., Green, W., Hamlet, A., Hinsley, W., Okell, L.C., van Elsland, S., … Ghani, A. C. (2020). Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf.

Tildesley, M. J., Vassall, A., Riley, S., Jit, M., Sandmann, F., Hill, E. M., Thompson, R. N., Atkins, B. D., Edmunds, J., Dyson, L., & Keeling, M. J. (2022). Optimal health and economic impact of non-pharmaceutical intervention measures prior and post vaccination in England: A mathematical modelling study. *Royal Society Open Science*, *9*(8), 211746. https://doi.org/10.1098/rsos.211746.

https://doi.org/10.1093/jrsssa/qnaf072
Advance access publication 26 June 2025



# Solveig Engebretsen, Alfonso Diz-Lois Palomares, Birgitte Freiesleben de Blasio, Arnoldo Frigessi, Geir Storvik, and Thordis L. Thorarinsdottir's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Solveig Engebretsen[1], Thordis L. Thorarinsdottir[2], Alfonso Diz-Lois Palomares[2,3], Geir Storvik[1,2], Arnoldo Frigessi[4,5] and Birgitte F. de Blasio[3,4]**

[1]SAMBA, Norwegian Computing Center, P.O. Box 114 Blindern, NO-0314 Oslo, Norway
[2]Department of Mathematics, University of Oslo, NO-0316 Oslo, Norway
[3]Division of Infection Control, Norwegian Institute of Public Health, N-0213 Oslo, Norway
[4]Oslo Centre for Biostatistics and Epidemiology, University of Oslo, 0317 Oslo, Norway
[5]Integreat, Norwegian Centre for Knowledge-driven Machine Learning, NO-0316 Oslo, Norway

*Address for correspondence*: Solveig Engebretsen, SAMBA, Norwegian Computing Center, P.O. Box 114 Blindern, NO-0314 Oslo, Norway. Email: solveig.engebretsen@nr.no

We commend the authors on discussing the modelling and statistical analyses, which is important for future pandemic preparedness. However, there is a crucial distinction between real-time operational context, amid an evolving crisis, and retrospective analyses. Health economics requires a broad set of underlying evaluations, which should be conducted in peacetime and not rushed during an unfolding pandemic. During the pandemic, modellers were responsible for performing high-quality research under immense time pressure, and communicating to policymakers and the public. Modelling choices and inference must be related to this working rhythm and the availability of data; researchers had a substantial impact on policymaking. Indeed, we were impressed by the timeliness of models like Flaxman et al. (2020).

As members of the modelling team of the Norwegian Institute of Public Health, we worked continuously on model improvements, incorporating dynamic epidemiological and environmental factors to address pressing questions.

Wood et al. highlight the lack of individual heterogeneity in COVID-19 models, focusing on contacts. Though transmission was characterised by heterogeneity in, e.g. contact structure and superspreading events, modelling individual heterogeneity requires detailed data and/or distributional assumptions, both difficult to obtain. Furthermore, contacts are not random, but tend to be regular (households, colleagues, etc.). In Norway, a key objective was prediction of short-term hospitalisation, an aspect only briefly considered by Wood et al. Here, heterogeneity would

require different hospitalisation risks depending on individual contact structure/infectivity/susceptibility. Furthermore, model-based short-term predictions with transmissibility parameters estimated from recent data, are likely less sensitive to heterogeneity assumptions.

There is no single best model for all questions (and timeframes for delivering answers). Like the UK, Norway operated with different models tailored to specific purposes. While our situational awareness-model (Engebretsen et al., 2023; Storvik et al., 2023) included regional and age-specific heterogeneity, significant heterogeneity was incorporated in our individual-based models used to explore strategies and scenarios (Chan et al., 2024; Ruscio et al., 2019). However, such models are challenging to calibrate.

Wood et al. criticise models for not including dynamic transmissibility. We believe that models with daily reproduction numbers (Storvik et al., 2023) are appropriate for monitoring changes in transmissibility, while we used less flexible step-functions for short-term predictions (Engebretsen et al., 2023).

Wood et al. criticise the use of test data without considering dynamic testing criteria. We agree, and indeed only used hospital data in the early pandemic, and later during the massive use of at-home self-tests (Engebretsen & Aldrin, 2024). The drawbacks were somewhat delayed detection of changes and higher estimated uncertainty (Engebretsen et al., 2023; Storvik et al., 2023). Here, uncertainty quantification is challenging, but not impossible (e.g. Rougier, 2007; Stephenson et al., 2012), as evidenced by work in climate science.

*Conflict of interests:* None declared.

## References

Chan L. Y. H., Rø G., Midtbø J. E., Di Ruscio F., Watle S. S. V., Juvet L. K., Littmann J., Aavitsland P., Nygård K. M., Berg A. S., Bukholm G., Kristoffersen A. B., Engø-Monsen K., Engebretsen S., Swanson D., Palomares A. D.-L., Lindstrøm J. C., Frigessi A., & de Blasio B. F. (2024). Modeling geographic vaccination strategies for COVID-19 in Norway. *PLoS Computational Biology*, *20*(1), 1–29. https://doi.org/10.1371/journal.pcbi.1011426

Engebretsen S., & Aldrin M. (2024). Effect of testing criteria for infectious disease surveillance: The case of COVID-19 in Norway. *PLoS One*, *19*(8), 1–14. https://doi.org/10.1371/journal.pone.0308978

Engebretsen S., Diz-Lois Palomares A., Rø G., Kristoffersen A. B., Lindstrøm J. C., Engø-Monsen K., Kamineni M., Hin Chan L. Y., Dale Ø., Midtbø J. E., Stenerud K. L., Di Ruscio F., White R., Frigessi A., & de Blasio B. F. (2023). A real-time regional model for COVID-19: Probabilistic situational awareness and forecasting. *PLoS Computational Biology*, *19*(1), 1–26. https://doi.org/10.1371/journal.pcbi.1010860

Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., Monod M., Ghani A. C., Donnelly C. A., Riley S., Vollmer M. A. C., Ferguson N. M., Okell L. C., & Bhatt S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Rougier J. (2007). Probabilistic inference for future climate using an ensemble of climate model evaluations. *Climatic Change*, *81*(3–4), 247–264. https://doi.org/10.1007/s10584-006-9156-9

Ruscio F. D., Guzzetta G., Bjørnholt J. V., Leegaard T. M., Moen A. E. F., Merler S., & de Blasio B. F. (2019). Quantifying the transmission dynamics of MRSA in the community and healthcare settings in a low-prevalence country. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(29), 14599–14605. https://doi.org/10.1073/pnas.1900959116

Stephenson D. B., Collins M., Rougier J. C., & Chandler R. E. (2012). Statistical problems in the probabilistic prediction of climate change. *Environmetrics*, *23*(5), 364–372. https://doi.org/10.1002/env.v23.5

Storvik G., Diz-Lois Palomares A., Engebretsen S., Rø G. Ø. I., Engø-Monsen K., Kristoffersen A. B., de Blasio B. F., & Frigessi A. (2023). A sequential Monte Carlo approach to estimate a time-varying reproduction number in infectious disease models: The Covid-19 case. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *186*(4), 616–632. https://doi.org/10.1093/jrsssa/qnad043

https://doi.org/10.1093/jrsssa/qnaf080
Advance access publication 27 June 2025

# Zhou Fang's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Zhou Fang**

Biomathematics and Statistics Scotland, JCMB, Kings Buildings, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK

*Address for correspondence:* Zhou Fang, Biomathematics and Statistics Scotland, JCMB, Kings Buildings, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK. Email: zhou.fang@bioss.ac.uk



The authors identify important shortcomings in the statistical work around the COVID-19 pandemic.

However, there are large uncertainties. For example, intense global discussion of lockdowns preceded their UK legal enactment (e.g. Google, 2020), allowing causally related effects before 23 March, even in countries without government response. UK public transport use already sharply declined a week before, and bus usage in Stockholm declined earlier than in London (Jenelius & Cebecauer, 2020; UK Government, 2025). Comparisons of Scandinavian countries, which may facilitate a more valid comparison than UK vs Sweden, could find no clear economic benefit from Sweden's approach to the pandemic (Blytt et al, 2022). Thus, a full accounting of what a lockdown actually means in terms of changes in behaviour and transmission, and short- and long-term economic impact is significantly more complex than the paper presents.

More specific to my own work, the paper criticizes case data as inferior to ONS CIS survey data. While there are valid caveats, there is also evidence in support of case data use. In Scotland, we analysed wastewater (WW) for changes in concentration of COVID-19 RNA from August 2020 (Fang et al, 2022). As shown in Figure 1, until the appearance of Omicron in December 2021, we saw an extremely strong relationship between WW RNA and case rates, including on the local authority level (after log transformation, $\rho = 0.97$ for national data, $\rho = 0.86$ for local authority data). The correlation appears to be stronger than with ONS CIS data ($\rho = 0.88$ in a similar period), with WW data showing similar patterns of a sharper rise during waves and tending to lead changes in CIS. This may reflect that both viral shedding into WW and cases are impacted by

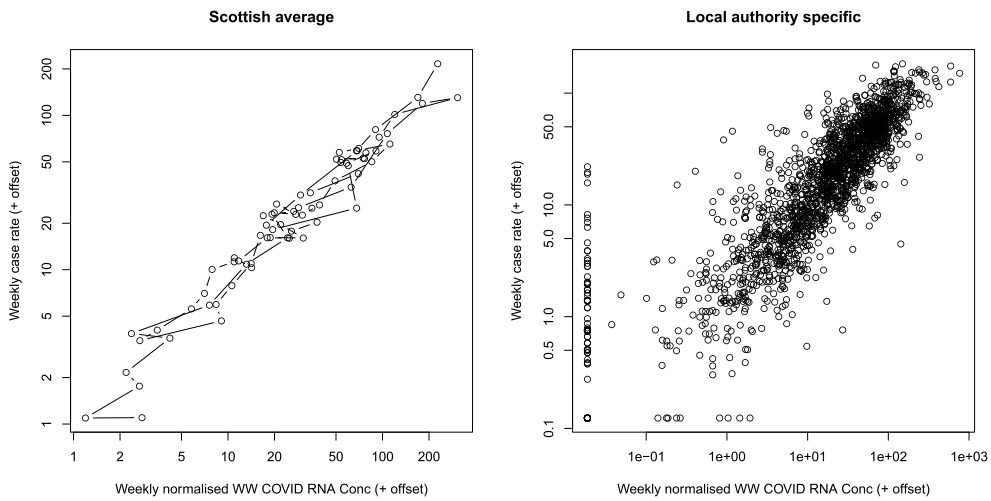



**Figure 1.** Scottish wastewater RNA concentrations vs case rates up until December 2021. A small offset is added to deal with zeroes.

disease severity, in which case the divergence between cases (and WW) and ONS CIS represents something real and important to pandemic response, not a shortcoming of the measure. Case data also offers greater rapidity and spatiotemporal resolution, which is very useful for calibration and for evaluating the effectiveness of interventions in real time.

Hence, I would argue that all these measures have their own uses, if used cautiously. While survey data does have sampling related advantages, if it is a lagging indicator that makes less distinction between serious and asymptomatic infection, it may not be the most informative and appropriate measure to present to the public.

*Conflicts of interest:* none declared.

## Data availability

Versions of the source data are publicly available. COVID-19 case data is from PHS (https://www.opendata.nhs.scot/dataset/covid-19-in-scotland) from which ongoing wastewater monitoring data is available. The specific processed data underlying this article will be shared on reasonable request to the corresponding author.

## References

Blytt, J. P., Bougroug, A., & Sletten, P. (2022). Economic development through Covid-19. An updated comparison of Norway, Sweden and Denmark. https://www.ssb.no/en/nasjonalregnskap-og-konjunkturer/konjunkturer/artikler/economic-development-through-covid-19.an-updated-comparison-of-norway-sweden-and-denmark (Accessed 2 July 2025).

Fang, Z., Roberts, A. M. I., Mayer, C.-D., Frantsuzova, A., Potts, J. M., Cameron, G. J., Singleton, P. T. R., & Currie, I. (2022). Wastewater monitoring of COVID-19: A perspective from Scotland. *Journal of Water and Health*, *20*(12), 1688–1700. https://doi.org/10.2166/wh.2022.082

Google. (2020). Google trends data on search interest in "lockdown". https://trends.google.com/trends/explore?date=2020-03-01%202020-04-15&q=lockdown (Accessed 2 July 2025).

Jenelius, E., & Cebecauer, M. (2020). Impacts of COVID-19 on public transport ridership in Sweden: Analysis of ticket validations, sales and passenger counts. *Transportation Research Interdisciplinary Perspectives*, *8*, 100242. https://doi.org/10.1016/J.TRIP.2020.100242

UK Government. (2025). Daily domestic transport use by mode. Official Statistics. https://www.gov.uk/government/statistics/transport-use-during-the-coronavirus-covid-19-pandemic (Accessed 2 July 2025).

https://doi.org/10.1093/jrsssa/qnaf089
Advance access publication 28 June 2025

# Thomas R. Fanshawe's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Thomas R. Fanshawe**[1]

[1]Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

*Address for correspondence*: Thomas R. Fanshawe, Nuffield Department of Primary Care Health Sciences, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK. Email: thomas.fanshawe@phc.ox.ac.uk

**Keywords:** long COVID, risk, epidemiology

My comments relate to Section 2.1 and the discussion of long COVID risk (LC).

Wood et al. claim 'a novel virus was always likely to result in an increase in people suffering longer term post-viral complications'. This surely is a case of being wise after the event. They state that public health messaging should have used the 'actual risk' but, confusingly, also that 'the long COVID evidence base was of insufficient statistical quality'. So what is this 'actual risk'? No one believes that the percentage of infections resulting in LC is as high as 45%, or as low as 0%. Estimates around 6% in adults are more realistic (Hanson et al., 2022). Nevertheless, this represents a huge cumulative health burden: LC can be a life-changing condition whose potential effects include damage to all organ systems (Al-Aly et al., 2024). But in 2020, this was still unknown and insufficient time had passed to quantify post-viral effects. Public health messaging exclusively emphasized mortality risk. We now know more about LC, yet Wood's discussion of the impact of lockdowns still focuses on mortality, without considering how many LC cases might have been prevented.

Inexplicably, Wood et al. then claim that exposure is associated with low/no difference in persistent symptoms. However, simply asserting that a study is 'carefully designed' does not make it so. One such study is a review whose own limitations state 'the majority of the included studies were of poor quality' (Behnood et al., 2022). Why not instead discuss the CloCk study? (Stephenson et al., 2023) This large, prospective cohort highlighted both elevated symptomatic risk post-PCR-confirmed infection, and also the characteristic relapsing/remitting pattern of these symptoms. They then give precedence to Matta et al. (2022), claiming 'subject blinding'; however, that study states 'participants were aware of their serology test results', and using serology as SARS-CoV-2 exposure is likely to cause appreciable diagnostic misclassification as many individuals do not seroconvert (Toh et al., 2022).

They also mention the 'problem' of the LC definition, which is necessarily broad enough to capture the bewildering variety of symptoms that patients experience. Many research studies already use tighter criteria, requiring e.g. a positive PCR result, but it is critically important not to require this of patients to access healthcare. For many people such access is already highly challenging (Greenhalgh et al., 2024). Patients should not be expected to prove exposure to the pathogen that caused their illness.

Ultimately, Wood et al. conclude 'none of this is to deny the existence of real sequelae to COVID infection'. Sadly, their unrepresentative and contradictory summary of the literature suggests that understating LC's risk and impact is precisely the impression they wish to convey.

*Conflicts of interest:* None declared.

## Data availability

No new data were generated or analysed in support of this research.

## References

Al-Aly, Z., Davis, H., McCorkell, L., Soares, L., Wulf-Hanson, S., Iwasaki, A., & Topol, E. J. (2024). Long COVID science, research and policy. *Nature Medicine*, *30*(8), 2148–2164. https://doi.org/10.1038/s41591-024-03173-6

Behnood, S. A., Shafran, R., Bennett, S. D., Zhang, A. X. D., O'Mahoney, L. L., Stephenson, T. J., Ladhani, S. N., De Stavola, B. L., Viner, R. M., & Swann, O. V. (2022). Persistent symptoms following SARS-CoV-2 infection amongst children and young people: A meta-analysis of controlled and uncontrolled studies. *Journal of Infection*, *84*(2), 158–170. https://doi.org/10.1016/j.jinf.2021.11.011

Greenhalgh, T., Sivan, M., Perlowski, A., & Nikolich, JŽ. (2024). Long COVID: A clinical update. *Lancet*, *404*(10453), 707–724. https://doi.org/10.1016/S0140-6736(24)01136-X

Hanson, S. W., Abbafati, C., Aerts, J. G., Al-Aly, Z., Ashbaugh, C., Ballouz, T., Blyuss, O., Bobkova, P., Bonsel, G., Borzakova, S., *et al.* (2022). Estimated global proportions of individuals with persistent fatigue, cognitive, and respiratory symptom clusters following symptomatic COVID-19 in 2020 and 2021. *JAMA*, *328*(16), 1604–1615. https://doi.org/10.1001/jama.2022.18931

Matta, J., Wiernik, E., Robineau, O., Carrat, F., Touvier, M., Severi, G., de Lamballerie, X., Blanché, H., Deleuze, J.-F., Gouraud, C., *et al.* (2022). Association of self-reported COVID-19 infection and SARS-CoV-2 serology test results with persistent physical symptoms among French adults during the COVID-19 pandemic. *JAMA Internal Medicine*, *182*(1), 19–25. https://doi.org/10.1001/jamainternmed.2021.6454

Stephenson, T., Pinto Pereira, S. M., Nugawela, M. D., McOwat, K., Simmons, R., Chalder, T., Ford, T., Heyman, I., Swann, O. V., Fox-Smith, L., *et al.* (2023). Long COVID—Six months of prospective follow-up of changes in symptom profiles of non-hospitalised children and young people after SARS-CoV-2 testing: A national matched cohort study (the CLoCk) study. *PLoS One*, *18*(3), e0277704. https://doi.org/10.1371/journal.pone.0277704

Toh, Z. Q., Anderson, J., Mazarakis, N., Neeland, M., Higgins, R. A., Rautenbacher, K., Dohle, K., Nguyen, J., Overmars, I., Donato, C., *et al.* (2022). Comparison of seroconversion in children and adults with mild COVID-19. *JAMA Network Open*, *5*(3), e221313. https://doi.org/10.1001/jamanetworkopen.2022.1313

https://doi.org/10.1093/jrsssa/qnaf077
Advance access publication 26 June 2025



# Jean Fisch's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Jean Fisch***

Management Consultant, Self-Employed, Brussels, Belgium

*Address for correspondence*: Email: fisch_jean@yahoo.com

The paper presents theses about 'questionable statistical methods or presentations' which do not seem to me to stand as they are.

## Thesis 1: Medical risk was over-represented

The appreciation whether any risk is misrepresented is 'in the eyes of the beholder.'

Any consideration of whether risks were over-represented can only be made objectively based on population level data (e.g. polls), not through personal considerations or comparisons.

More widely, the topic of medical risk being overrepresented looks more political/ethical (choosing between conflicting objectives) than statistical.

## Thesis 2: Excess deaths were exaggerated

The method used in the paper to estimate UK expected (and excess) deaths

- assumes 2020–2022 mortality rates to remain at 2017–2019 average levels, without further mortality gain
- offers no explanation why the 0.5% annual mortality gain trend since 2011 would suddenly stop in 2020

CMI, the monitoring mortality body of the Society of Actuaries, does assume some continued mortality gains based on long-term trends and estimates excess at ~150,000 (CMI Appendix, 1 of the Mortality Monitor – week, 1 of 2024, January, 2024), which is far above the 95,000 in the paper (and not far off ONS).

The critique of the ONS approach is valid but, at this stage, the proposed method/conclusions by the authors also raise questions.

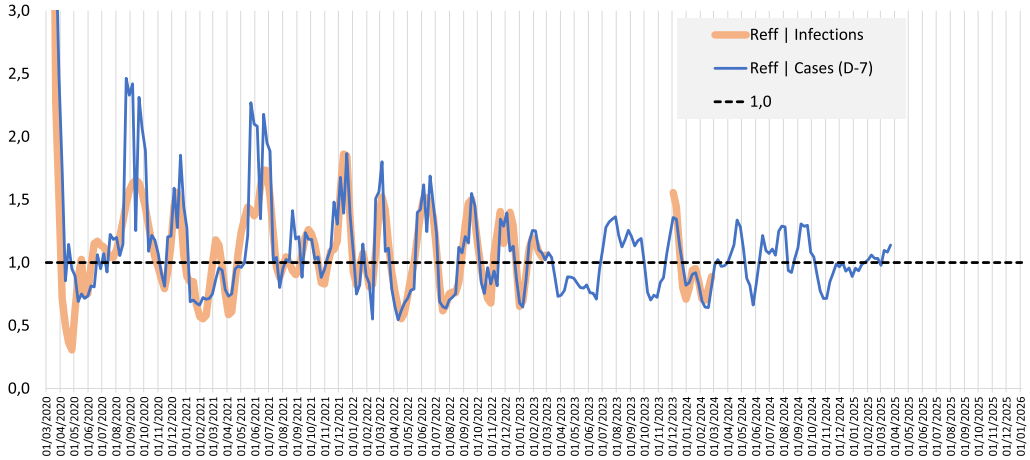



**Figure 1.** Comparison of "R Eff" derived from infections per ONS (Cases per UKHSA covid dashboard) and cases per UKHSA Infections per weekly 'ONS Infection Surveys' report (2023) for England.


**Figure 2.** COVID Measures introduced by the Federal Swiss Government in January 2021. Source: Switzerland announces lockdown–Swiss Radio & Television article in German, January (2021).

## Thesis 3: Cases misrepresent infections

The authors share a chart of 'ONS infection survey prevalence vs. reported cases' which shows some decorrelations.

However, a comparison of cases (Cases per UKHSA covid dashboard ) and infections evolution as estimated by ONS (Infections per weekly 'ONS Infection Surveys' report, 2023) (see Figure 1) shows that cases proved to be, by enlarge, a good proxy for the 'Reff' of infections with two exceptions (October 2020 and June 2021).

The critique of the authors that 'cases misrepresented the situation' does not seem justified: On the contrary, cases proved a (surprisingly) good early indicator of infections.

## Thesis 4: The necessity of lockdowns to cut transmission does not match the evidence

The authors claim: 'Sweden and Switzerland show that with much less measures, a similar objective could be achieved.'

The authors claim that Switzerland saw measures which were less harsh than those of its neighbours but this is not matched by the reports from the Swiss authorities of January 2021: The measures were similarly harsh as that of its neighbours. Switzerland opted for a stringent lockdown, with only

schools open (Switzerland announces lockdown—Swiss Radio & Television article in German, January, 2021) (as in Belgium and France) from mid-January 2021 to March 2021 (see Figure 2).

The suggestion put forward by the authors that there could have been a path without lockdown relies on the example of Sweden, for which the question of transferability to other countries is not addressed in this paper (nor, in fairness to the authors, anywhere else to the best of my knowledge).

*Conflicts of interest:* none declared.

## References

Cases per UKHSA covid dashboard. http://coronavirus.data.gov.uk. Date accessed December 15, 2023.

CMI Appendix 1 of the Mortality Monitor—week 1 of 2024 (January 2024): https://www.actuaries.org.uk/system/files/field/document/Mortality%20monitor%20Week%202024%–01%20v01%202024-01-17.pdf

Infections per weekly "ONS Infection Surveys" report (up to March 2023): https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases?page=1. Date accessed June 04, 2025.

"Switzerland announces lockdown"—Swiss Radio & Television article in German (January 2021): https://www.srf.ch/news/schweiz/verschaerfte-massnahmen-mit-dem-shutdown-droht-die-soziale-isolation. Date accessed June 04, 2025.

https://doi.org/10.1093/jrsssa/qnaf070
Advance access publication 26 June 2025

# Professor Ian Hall's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Ian Hall**

Professor of Mathematical Epidemiology and Statistics, University of Manchester, Manchester, UK

*Address for correspondence*: Ian Hall, Professor of Mathematical Epidemiology and Statistics, University of Manchester, Alan Turing Building, Manchester M13 9PL, UK. Email: Ian.hall@manchester.ac.uk

1. **Presentation of risk:** Young people will likely have grandparents who are at risk. This is the key part of the phrasing in *the adverts*. Total societal risk is then not necessarily just the sum of personal risks given the chain of transmissions, and whilst shielding was mooted, it would not effectively work in some settings (such as care home). Thus, a societal perspective is needed including wider well-being around family caring rather than a simple focus on personal risk of infection.
2. **Excess deaths**. In Morciano et al. (2020), the critical issue was that excess deaths were the only way to identify the missed cases/burden of disease from lack of testing/reliance on symptomatic cases in wave 1. We stopped using excess deaths as a metric in care homes after wave 1 due to 'harvesting' and replenishment of population affecting the 'baseline' (hence moving instead to Overton et al. 2021) when advising on severity changes to SAGE and subgroups.
3. **Missed cases:** Given cases were missed in wave 1 can we be sure studies capture true signal in wave 1. Even well-designed studies like React and ONS may have been riding the wave rather than tracking it until testing capacity was established.
4. **Smoother impact:** Figure 4 in Pang et al. (2025) shows the issue of a simulated step change in $R$ forcing the GAM to smooth trend, so there would be inferred modelled peak before intervention because of the smoother adapted given the choice of knot, which whilst not identical

to this scenario likely translates to such modelled states. Rather than retrospective to whole time series (and leveraging future information) one should run the model with data only up to the lockdown/intervention to see if the change is apparent—as we sort of do in Figure 4d [code here: doi:10.5281/zenodo.14045216].

5. **Infection time to death distributions.** Jamieson et al. (2024) looks at delays for incubation period but ideas would carry over to other delay distributions, in particular the censoring to whole day. The paragraph at top of p. 19 does not make it clear this is factored in. A 20 day delay is less sensitive to that than 5 day one but it may be a factor as an uncensored fit would be a day longer on average.
6. **Spontaneous behaviour change:** *if* model gives a true signal there may have been informal change before lockdown marginally reducing $R$, which maybe insufficient to bring $R < 1$.

*Conflicts of interest:* none declared.

## Funding

I.H. received funding from Wellcome Trust and ESPRC.

## References

Jamieson, N., Charalambous, C., Schultz, D. M., & Hall, I. (2024). The Burr distribution as a model for the delay between key events in an individual's infection history. *PLoS Computational Biology*, *20*, e1012041. https://doi.org/10.1371/journal.pcbi.1012041

Morciano, M., Stokes, J., Kontopantelis, E., Hall, I., & Turner, A. J. (2020). Excess mortality for care home residents during the first 23 weeks of the COVID-19 pandemic in England: A national cohort study. *BMC Medicine*, *19*, 71. https://doi.org/10.1186/s12916-021-01945-2

Overton, C. E., Webb, L., Datta, U., Fursman, M., Hardstaff, J., Hiironen, I., Paranthaman, K., Riley, H., Sedgwick, J., Verne, J., Willner, S., Pellis, L., & Hall, I. (2021). Novel methods for estimating the instantaneous and overall COVID-19 case fatality risk among care home residents in England. *PLoS Computational Biology*, *18*, e1010554. https://doi.org/10.1371/journal.pcbi.1010554

Pang, X., Han, Y., Tressier, E., Abdul Aziz, N., Pellis, L., House, T., & Hall, I. (2025). Time-varying reproduction number estimation: Fusing compartmental models with generalized additive models. *Journal of the Royal Society, Interface*, *22*, 20240518. https://doi.org/10.1098/rsif.2024.0518

https://doi.org/10.1093/jrsssa/qnaf076
Advance access publication 27 June 2025

# Leonhard Held and Maria Bekker-Nielsen Dunbar's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Leonhard Held[1,2] and Maria Bekker-Nielsen Dunbar[3]**

[1]Center for Reproducible Science, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland
[2]Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland
[3]Department of Infectious Disease and Tropical Medicine, University Hospital Heidelberg, Im Neuenheimer Feld 324, 69120 Heidelberg, Germany

*Address for correspondence*: Leonhard Held, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland. Email: leonhard.held@uzh.ch

Wood et al. (2025) discuss statistical aspects of the COVID-19 response in the UK. We commend their work. We agree that the availability of data is key to strengthening trust in policy decisions during a pandemic (Schwab & Held, 2020; van Kleef et al., 2025). Needless to say, open data are only one of many aspects of reproducible epidemiologic research (Mathur & Fox, 2023; Bekker-Nielsen Dunbar & Held, 2024b; Fraga-González et al., 2025).

In our comment, we focus on statistical methods relating nonpharmaceutical interventions (NPIs) with disease incidence as their use may explain part of differences in incidence patterns between countries. Wood et al. (2025) infer COVID-19 incidence trajectories consistent with observed mortality in European countries using a deconvolution model and informally relate it to the timing of lockdowns, an NPI. Voluntary avoidance behaviours may further reduce incidence (Brankston et al., 2024), but are difficult to measure (Bish & Michie, 2010).

We used country-level information on NPIs during the COVID-19 pandemic (from Hale et al., 2021) to model the effect of school closures on the spread of COVID-19 in Switzerland (Bekker-Nielsen Dunbar et al., 2024). This was done using stratified endemic–epidemic regression models, providing detailed analysis of disease spread across age-groups and regions. Additional data on contacts between age-groups and mobility data for spatial dispersal were incorporated and adjustments for changes in incidence on weekdays and holidays were made. As in Bekker-Nielsen Dunbar and Held (2024a), a counterfactual approach was then applied to predict and compare the disease incidence in scenarios of different intervention policies.

Another aspect—less relevant for the UK, but important for landlocked countries—is the disease spread across countries. We investigated the effect of border closure on the initial spread of the disease (Grimée et al., 2022). This was an important question early on, since the first COVID-19 cases in Switzerland were imported from neighbouring Italy. Spatiotemporal endemic-epidemic models can provide more detailed insights than a simple time series analysis of aggregated death counts, where fine-scale regional differences may be overlooked.

Cross-country comparisons are easier to do with standardized death rates; mortality is frequently used to compare the impact of pandemics. An important issue raised by Wood et al. (2025) is the problem of inferring incidence or prevalence from case data, subject to underreporting and reporting delays (Noufaily, 2019). A limitation of our approach is its reliance on case data. If underreporting rates are time-constant, adjustments in the univariate endemic–epidemic model are available (Bracher & Held, 2021). If underreporting rates vary across time, age-group or region, inferring cases becomes problematic.

To capture underreporting, we included the testing rate as covariate in endemic–epidemic models for COVID-19 case data (Grimée et al., 2022) as testing has an inverse relationship with underreporting. This captures differences in testing strategies linked to the choice of NPIs (health status certificates vs. stay-at-home orders). The use of COVID-19 death data on such a fine level of analysis is prohibitive due to small counts (mostly zero) in younger age-groups and smaller regions. Later developments have addressed excess zeros in endemic–epidemic models (Lu & Meyer, 2023), but still require at least some events in all strata of age-groups and regions.

*Conflicts of interest:* None declared.

## References

Bekker-Nielsen Dunbar M., & Held L. (2024a). The COVID-19 vaccination campaign in Switzerland and its impact on disease spread. *Epidemics*, *47*(Supplement_1), 100745. https://doi.org/10.1016/j.epidem.2024.100745

Bekker-Nielsen Dunbar M., & Held L. (2024b). Improving reproducibility in epidemiology. In W. Ahrens, & I. Pigeot (Eds.), *Handbook of Epidemiology*. Springer.

Bekker-Nielsen Dunbar M., Hofmann F., Meyer S., & Held L. (2024). Endemic-epidemic modelling of school closure to prevent spread of COVID-19 in Switzerland. *BMC Infectious Diseases*, *24*(1), 1372. https://doi.org/10.1186/s12879-024-09674-6

Bish A., & Michie S. (2010). Demographic and attitudinal determinants of protective behaviours during a pandemic: A review. *British Journal of Health Psychology*, *15*(4), 797–824. https://doi.org/10.1348/135910710X485826

Bracher J., & Held L. (2021). A marginal moment matching approach for fitting endemic-epidemic models to underreported disease surveillance counts. *Biometrics*, *77*(4), 1202–1214. https://doi.org/10.1111/biom.v77.4

Brankston G., Fisman D. N., Poljak Z., Tuite A. R., & Greer A. L. (2024). Examining the effects of voluntary avoidance behaviour and policy-mediated behaviour change on the dynamics of SARS-CoV-2: A mathematical model. *Infectious Disease Modelling*, *9*(3), 701–712. https://doi.org/10.1016/j.idm.2024.04.001

Fraga-González G., van de Wiel H., Garassino F., Kuo W., de Zélicourt D., Kurtcuoglu V., Held L., & Furrer E. (2025). Affording reusable data: Recommendations for researchers from a data-intensive project. *Scientific Data*, *12*(1), 258. https://doi.org/10.1038/s41597-025-04565-0

Grimée M., Bekker-Nielsen Dunbar M., Hofmann F., & Held L. (2022). Modelling the effect of a border closure between Switzerland and Italy on the spatiotemporal spread of COVID-19 in Switzerland. *Spatial Statistics*, *49*(5), 100552. https://doi.org/10.1016/j.spasta.2021.100552

Hale T., Angrist N., Goldszmidt R., Kira B., Petherick A., Phillips T., Webster S., Cameron-Blake E., Hallas L., Majumdar S., & Tatlow H. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, *5*(4), 529–538. https://doi.org/10.1038/s41562-021-01079-8

Lu J., & Meyer S. (2023). A zero-inflated endemic-epidemic model with an application to measles time series in Germany. *Biometrical Journal*, *65*(8), e2100408. https://doi.org/10.1002/bimj.202100408

Mathur M. B., & Fox M. P. (2023). Toward open and reproducible epidemiology. *American Journal of Epidemiology*, *192*(4), 658–664. https://doi.org/10.1093/aje/kwad007

Noufaily A. (2019). Underreporting and reporting delays. In L. Held, N. Hens, P. O'Neill, & J. Wallinga (Eds.), *Handbook of infectious disease data analysis* Ch. 22. Chapman and Hall/CRC.

Schwab S., & Held L. (2020). Science after Covid-19: Faster, better, stronger? *Significance*, *17*(4), 8–9. https://doi.org/10.1111/1740-9713.01415

van Kleef E., Van Bortel K., Arsevska E., Busani L., Dellicour S., Di Domenico L., Gilbert M., van Elsland S. L., Kraemer M. U. G., Lai S., Lemey P., Merler S., Milosavljevic Z., Rizzoli A., Simic D., Tatem A. J., Maguelonne Teisseire M., Wint W., Colizza V., & Poletto C. *Modelling practices, data provisioning, sharing and dissemination needs for pandemic decision-making: A European survey-based modellers' perspective*. medRxiv. https://doi.org/10.1101/2025.03.12.25323819, 2025, preprint: not peer reviewed.

Wood S. N., Wit E. C., McKeigue P. M., Hu D., Flood B., Corcoran L., & Jawad T. A. (2025). Some statistical aspects of the Covid-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society*. *49*.

https://doi.org/10.1093/jrsssa/qnaf106
Advance access publication 4 July 2025

# Edward Hill's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Edward M. Hill**

Civic Health Innovation Labs and Institute of Population Health, University of Liverpool, Liverpool, UK

*Address for correspondence*: Edward M. Hill, Civic Health Innovation Labs, Liverpool Science Park, 131 Mount Pleasant, Liverpool L3 5TF, UK. Email: Edward.Hill@liverpool.ac.uk

I provide here perspective comments on Section 3—'Covid cases and other media distortions'. I agree that cases often provide a biased picture of incidence or prevalence. However, they were a simple measure that had been available for much of the pandemic that could be used to convey the general trends in infection. Whether the government reporting should have switched to a new metric part-way through the pandemic is a question for those who specialize in communication. I note that the scaling between Office for National Statistics (ONS) estimates and cases used in this article is a subjective choice. The authors have chosen to match close to a minimum (around March 2021), which naturally leads to a greater deviation near the peaks. While the pandemic was ongoing, however, many sources of data were used by advisors and ministers to make policy decisions, including case and testing data,

hospitalization data, deaths data, ONS and REACT survey data and others. Each data source has strengths and weaknesses, particularly in terms of their reliability, precision, bias, and how far they lag behind new infections. The media also reported multiple data sources, with BBC news in March 2021, for example, including cases, deaths, people in hospital and vaccinations. Building a coherent pattern from these multiple complex data sources remains an open statistical challenge.

I agree with the authors that COVID-19 deaths need to be placed into context. Comparison to bubonic plague in the Middle Ages (when life expectancy was around 30 years and childhood mortality extremely high) might not be the most relevant to today's population. In 2020 and 2021, the ONS reports that COVID-19 was the leading cause of death, although it dropped to sixth in 2022 (Office for National Statistics, 2023a); in 2010, it was estimated that 7% of all deaths (approximately 40,000) were due to infectious diseases (Houses of Parliament Parliamentary Office of Science and Technology, 2017; Office for National Statistics, 2023b)—the death toll from COVID-19 in 2020 far exceeds this value.

*Conflicts of interest:* none declared.

## References

Houses of Parliament Parliamentary Office of Science and Technology. (2017). *UK Trends in Infectious Disease*. http://researchbriefings.files.parliament.uk/documents/POST-PN-0545/POST-PN-0545.pdf. Date accessed April 7, 2025.

Office for National Statistics. (2023a). *Deaths registered in England and Wales: 2022*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregistrationsummarytables/2022. Date accessed April 7, 2025.

Office for National Statistics. (2023b). *Vital statistics in the UK: births, deaths and marriages*. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/vitalstatisticspopulationandhealthreferencetables. Date accessed April 7, 2025.

https://doi.org/10.1093/jrsssa/qnaf071
Advance access publication 1 July 2025

# Thomas House's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Thomas House**

Department of Mathematics, University of Manchester, Manchester M13 9PL, UK

*Address for correspondence*: Thomas House, Department of Mathematics, University of Manchester, Manchester M13 9PL, UK. Email: thomas.house@manchester.ac.uk

During the COVID-19 pandemic, I held a visiting position at the local NHS to collaborate on quantitative assessment of hospital burden; our preliminary work looking at hospital surges actually involved the heterogeneous susceptibility model of Novozhilov (2008); however, it quickly became clear that the level of heterogeneity would have to be extremely high to make any material difference during the first wave. This is in fact shown in Figure 8 of the paper by Wood et al. For the situation where $\mathcal{R}_0 = 2$ and the 'immunity coefficient' $\lambda = 5$, there is still an infection probability of 20%, which is well in excess of the antibody prevalence of 6% estimated for mid-July 2020 by the REACT-2 study (Ward et al., 2021).

For the 2009–2010 H1N1 influenza pandemic, the total percentages experiencing infection are believed to be much lower than the estimates of $\mathcal{R}_0$ would suggest (Van Kerkhove et al., 2013), likely due to pre-existing immunity to related viruses, and the modelling community was and remains well aware of this. Wood et al. are incorrect that 'this effect was neglected', for example, a 2020 version of the work of Gomes et al. (2022) was commented on by Graham Medley, at the time chair of SPI-MO (Medley, 2020).

Elsewhere, Britton et al. (2020) argued that herd immunity thresholds might be around two thirds of that in the simplest models. This was done using realistic age and household structure rather than a generic susceptibility distribution but would seem to correspond to values of $\lambda$ below 2.5 in Figure 8 of Wood et al. Other work by Fox et al. (2020) took the approach of Gomes et al. (2022) seriously but drew different conclusions about likely values of model parameters. The implication from Wood et al. that 'models calibrated against data' all point to values of $\lambda$ above 2.5 is therefore somewhat misleading.

The final misleading claim that I would like to address from Wood et al. is that heterogeneity has 'a large unidirectional effect on results'. This is not correct, for example, when the division of the

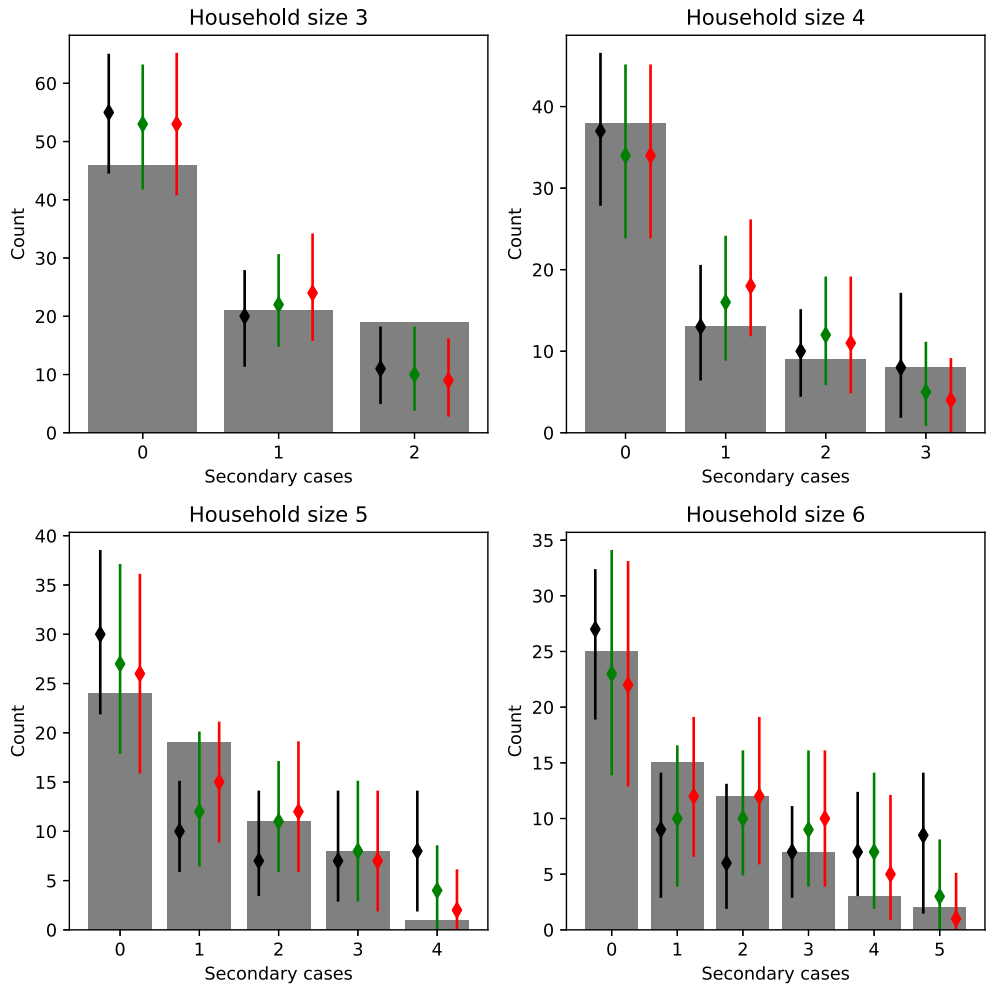



**Figure 1.** Household data from Dattner et al. (2021) (grey bars) is shown against fitted models that are the individual-based equivalents of the variable susceptibility model of Novozhilov (2008) discussed in Section 4.1 of Wood et al. 95% posterior predictive intervals are shown for 'immunity coefficients' $\lambda = 1.2$ (left, black), $\lambda = 2.9$ (middle, green), $\lambda = 4.7$ (right, red), quoted in Section 4.1 of Wood et al.

population into households is accounted for, or there is otherwise significant transitivity in the contact network, since then we can see larger final sizes than would be expected from simpler models (House & Keeling, 2011; Pellis et al., 2020).

Wood et al. do offer the constructive challenge that maybe 'it was felt to be difficult or impossible to estimate the person to person variability in transmission rates'. I believe that household data does offer a route to such estimation; for example Tsang et al. (2023) used such data to estimate heterogeneity in infectivity; there is no *a priori* reason that this should not be extended to estimation of the models considered in Section 4.1 of Wood et al.

To consider this, I used a subset of household data from Dattner et al. (2021), and Approximate Bayesian Computation for household data (Toni & Stumpf, 2009), adapting the analysis of House (2014) to derive a model with tuneable susceptibility. Code and extra analysis is available at https://github.com/thomasallanhouse/covid19-incidence/blob/main/Het.ipynb.

This yields the results shown in Figure 1. Clearly, these are not decisive about the level of heterogeneity implied by early household data but equally the histograms imply potential identifiability, which may be possible if multiple datasets are integrated to obtain more statistical evidence as in Tsang et al. (2023). My personal belief is that paying attention to such detailed questions is likely to have much more benefit for pandemic preparedness than sweeping and poorly evidenced criticisms of the large and diverse expert community—containing many statisticians—of quantitative scientists who contributed to the COVID-19 response.

*Conflict of interests:* None declared.

## References


Britton T., Ball F., & Trapman P. (2020). A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science*, *369*(6505), 846–849.

Dattner I., Goldberg Y., Katriel G., Yaari R., Gal N., Miron Y., Ziv A., Sheffer R., Hamo Y., & Huppert A. (2021). The role of children in the spread of COVID-19: Using household data from Bnei Brak, Israel, to estimate the relative susceptibility and infectivity of children. *PLOS Computational Biology*, *17*(2), e1008559.

Fox S. J., Potu P., Lachmann M., Srinivasan R., & Meyers L. A. (2020). *The COVID-19 herd immunity threshold is not low: A re-analysis of European data from spring of 2020*. medRxiv. https://doi.org/10.1101/2020.12.01.20242289, preprint: not peer reviewed. https://www.medrxiv.org/content/early/2020/12/03/2020.12.01.20242289

Gomes M. G. M., Ferreira M. U., Corder R. M., King J. G., Souto-Maior C., Penha-Gonçalves C., Gonçalves G., Chikina M., Pegden W., & Aguas R. (2022). Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold. *Journal of Theoretical Biology*, *540*, 111063.

House T. (2014). Non-Markovian stochastic epidemics in extremely heterogeneous populations. *Mathematical Modelling of Natural Phenomena*, *9*(2), 153–160.

House T., & Keeling M. J. (2011). Epidemic prediction and control in clustered populations. *Journal of Theoretical Biology*, *272*(1), 1–7.

Medley G. (2020). I wondered what Gabriela was doing… she has a better intuitive understanding of heterogeneity and dynamics than any I have met. I will read with interest. https://twitter.com/GrahamMedley/status/1257065945345986560 Posted 10:54 PM May 3, from Leamington Spa, England.

Novozhilov A. S. (2008). On the spread of epidemics in a closed heterogeneous population. *Mathematical Biosciences*, *215*(2), 177–185.

Pellis L., Cauchemez S., Ferguson N. M., & Fraser C. (2020). Systematic selection between age and household structure for models aimed at emerging epidemic predictions. *Nature Communications*, *11*, 906.

Toni T., & Stumpf M. P. H. (2009). Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics*, *26*(1), 104–110.

Tsang T. K., Huang X., Wang C., Chen S., Yang B., Cauchemez S., & Cowling B. J. (2023). The effect of variation of individual infectiousness on SARS-CoV-2 transmission in households. *eLife*, *12*, e82611.

Van Kerkhove M. D., Hirve S., Koukounari A., Mounts A. W., & for the H1N1pdm serology working group (2013). Estimating age-specific cumulative incidence for the 2009 influenza pandemic: a meta-analysis of A(H1N1)pdm09 serological studies from 19 countries. *Influenza and Other Respiratory Viruses*, *7*(5), 872–886.

Ward H., Atchison C., Whitaker M., Ainslie K. E. C., Elliott J., Okell L., Redd R., Ashby D., Donnelly C. A., Barclay W., Darzi A., Cooke G., Riley S., & Elliott P. (2021). SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nature Communications*, *12*, 905.

https://doi.org/10.1093/jrsssa/qnaf115
Advance access publication 8 July 2025

# John Ioannidis' contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**John P.A. Ioannidis**

Departments of Medicine, of Epidemiology and Population Health, and of Biomedical Data Science and Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA, USA

*Address for correspondence:* John P.A. Ioannidis, 3180 Porter Dr, Stanford Research Park, Rm A129, Palo Alto, CA 94304, USA. Email: jioannid@stanford.edu

Wood et al. dissect several important aspects surrounding the COVID-19 response that reflect overestimation of COVID-19 risk and underestimation of risks from the pandemic response. Cumulative biases of large magnitude have created a misleading pandemic narrative. Defending this implausible narrative at all costs has requested continuous introduction of yet new biases, new extreme estimates, and even more implausible follow-up patchwork. All this in the hope that these spurious reinforcements would delay the narrative's eventual collapse. A massive and obviously flawed literature of opinions (Kepp et al., 2025) and modelling (Chin et al., 2021; Ioannidis et al., 2022; Mueller et al., 2025) is impossible to defend currently in scientific terms. It is simply maintained by the gravitas of policy-makers, influencers, media, journal editors, and some scientists who remain unwilling to accept that they might have been so devastatingly wrong. Leading journals are littered with hundreds of papers for which even retraction would not be sufficient to correct the record, since they have already caused tremendous damage.

The way that science was hijacked by advocacy, activism, and conflicted stakeholders during the COVID-19 pandemic will require careful study for many years to come. This forensic meta-research work will probably require a combination of many types of expertise, ranging from statistics to psychology and sociology. It may also require a greater time distance, when the perpetrators will no longer be able to smear and silence opposition or divert comments away from the main issues.

Currently, one can only say that at a global level, the damage done by the measures adopted probably caused far more harm and more deaths than SARS-CoV-2 itself; that most projected or imputed estimates of life-years lost due to SARS-CoV-2 infections were exaggerated; and that most aggressive policies adopted had major blind spots and disregard for accumulating evidence. Lockdown measures were unlikely to reduce infections much and save lives in the long term once the virus had been introduced to a community and epidemic waves had started (Bendavid & Patel, 2024). They were almost certain to cause major harms, and they unfortunately did, leading also to a death spiral of faulty decision-making (Schippers et al., 2024). Faulty decision-making resulted in loss of trust in science and offered vast ammunition to weird conspiracy theories. The eventual battle between pandemic zealots and denial conspirators has left evidence-based science torn to pieces between belligerent tribes.

*Conflicts of interest:* none declared.

## References

Bendavid, E., & Patel, C. J. (2024). Epidemic outcomes following government responses to COVID-19: Insights from nearly 100,000 models. *Science Advances*, *10*(23), eadn0671. https://doi.org/10.1126/sciadv.adn0671

Chin, V., Ioannidis, J. P. A., Tanner, M. A., & Cripps, S. (2021). Effect estimates of COVID-19 non-pharmaceutical interventions are non-robust and highly model-dependent. *Journal of Clinical Epidemiology*, *136*, 96–132. https://doi.org/10.1016/j.jclinepi.2021.03.014

Ioannidis, J. P. A., Cripps, S., & Tanner, M. A. (2022). Forecasting for COVID-19 has failed. *International Journal of Forecasting*, *38*(2), 423–438. https://doi.org/10.1016/j.ijforecast.2020.08.004

Kepp, K. P., Cristea, I., Muka, T., & Ioannidis, J. P. A. (2025). COVID-19 advocacy bias in the BMJ: Meta-research evaluation. *BMJ Open Quality*, *14*(1), e003131. https://doi.org/10.1136/bmjoq-2024-003131

Mueller, B., Padberg, I., Lorke, M., Brinks, R., Cripps, S. Gomes M.G.W., Haake D., & Ioannidis J.P. *Uncertainty and inconsistency of COVID-19 non-pharmaceutical intervention effects with multiple competitive statistical models*. medRxiv https://doi.org/10.1101/2025.01.22.25320783, 25 January, 2025, preprint: not peer reviewed.

Schippers, M. C., Ioannidis, J. P. A., & Luijks, M. W. J. (2024). Is society caught up in a Death Spiral? Modeling societal demise and its reversal. *Frontiers in Sociology*, *9*, 1194597. https://doi.org/10.3389/fsoc.2024.1194597

Wood, S., Wit, E. C., McKeigue, P. M., Hu, D., Flood, B., Corcoran, L., & Jawad, T. A. Some statistical aspects of the COVID-19 response. *Journal of the Royal Statistical Society Series A*, this issue.

https://doi.org/10.1093/jrsssa/qnaf079
Advance access publication 30 June 2025

# Oliver Johnson's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Oliver Johnson**

School of Mathematics, University of Bristol, Fry Building, Woodland Road, Bristol BS8 1UG, UK

*Address for correspondence*: Oliver Johnson, School of Mathematics, University of Bristol, Fry Building, Woodland Road, Bristol BS8 1UG, UK. Email: O.Johnson@bristol.ac.uk

As a data-driven 'COVID centrist', conscious of the harms of both disease and NPIs, I welcome this paper Wood et al. (2025) and other attempts to calibrate the effectiveness of the UK response. However, I would like to raise some points:

## 1 Section 3: Cases and ONS

### 1.1 Cases and prevalence

While of course cases (positive tests) are not the same as prevalence (total infected people), I believe most media and commentators generally made that distinction clear. As an illustration, BBC graphs such as Figure 1 typically (if somewhat arbitrarily) distinguished 'targeted testing' and 'mass testing' regimes.

The 'curious argument' (Wood et al., 2025, P. 12) that short-term fluctuations in test numbers can be ignored to first approximation can be somewhat justified by plotting week-on-week ratios of test numbers (UK Health Security Agency (UKHSA), 2022b) (Figure 2). This shows that typical weekly changes in test numbers were on the order of 10% or less, much smaller in magnitude than

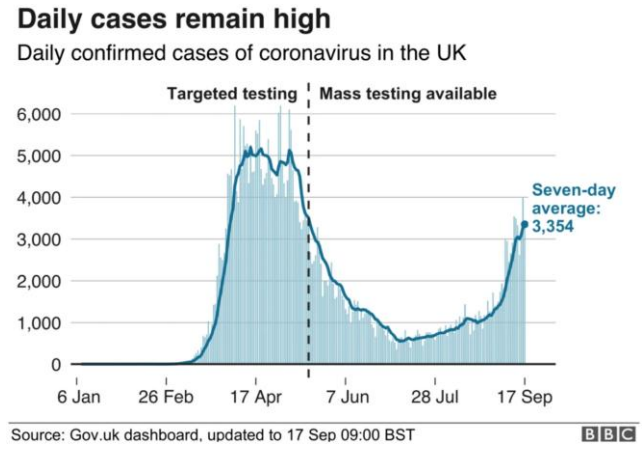



**Figure 1.** Illustration of typical BBC graph of case data, taken from 18th September 2020 (BBC News, 2020).

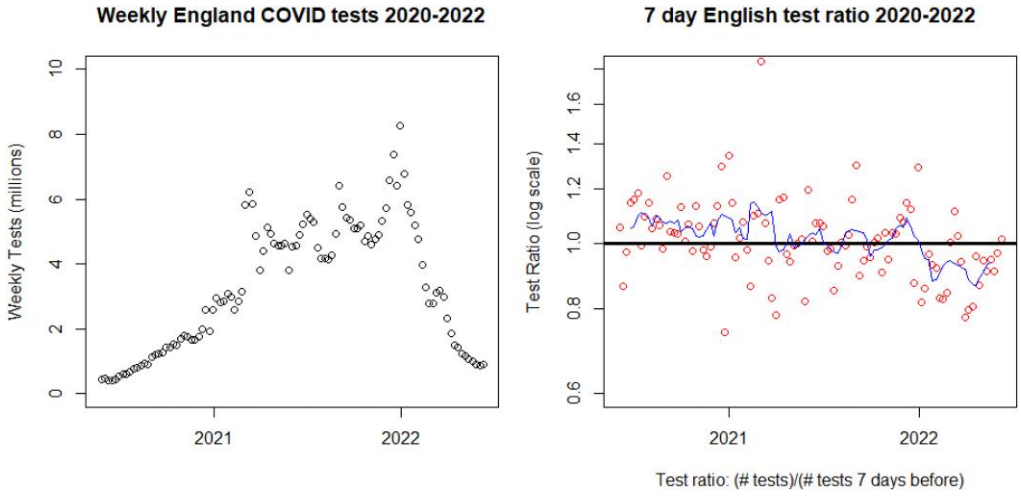



**Figure 2.** Data on weekly tests in England (28th May 2020–9th June 2022) from UK Health Security Agency (UKHSA) (2022b). a) Absolute weekly numbers (linear scale). b) Ratio of number of tests in a given week to number in previous week. Dots represent weekly change, line is smoothed average of three points. The single week with very high ratio (big increase in testing) corresponds to introduction of lateral flow tests in schools for March 2021 reopening.

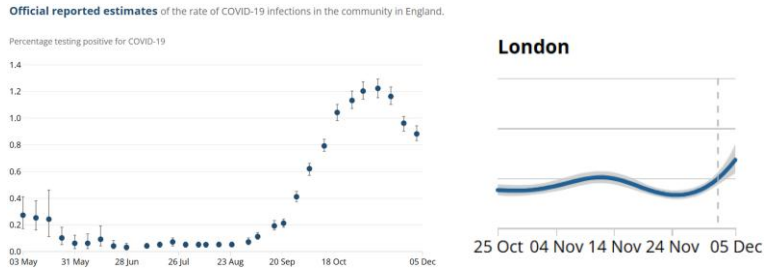



**Figure 3.** Graphs of estimated COVID prevalence taken from ONS survey, 11th December 2020; a) England and b) London.

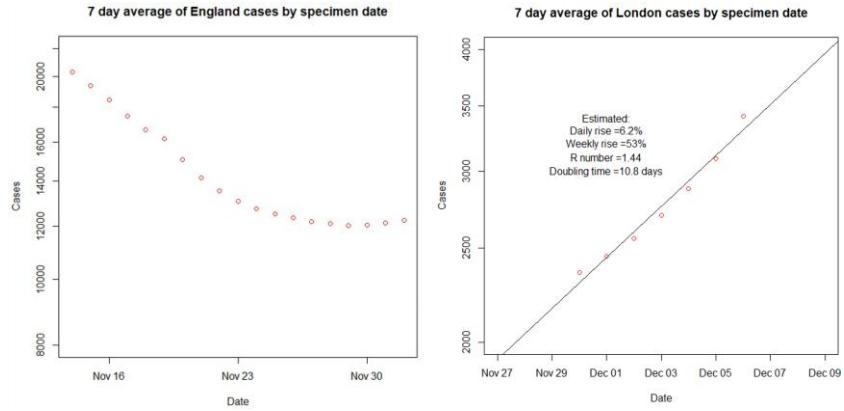



**Figure 4.** Graphs tweeted by Oliver Johnson @BristOliver based on contemporaneous case data; a) 9th December 2020 (England) Johnson (2020a) and b) 14th December 2020 (London) (Johnson, 2020b).

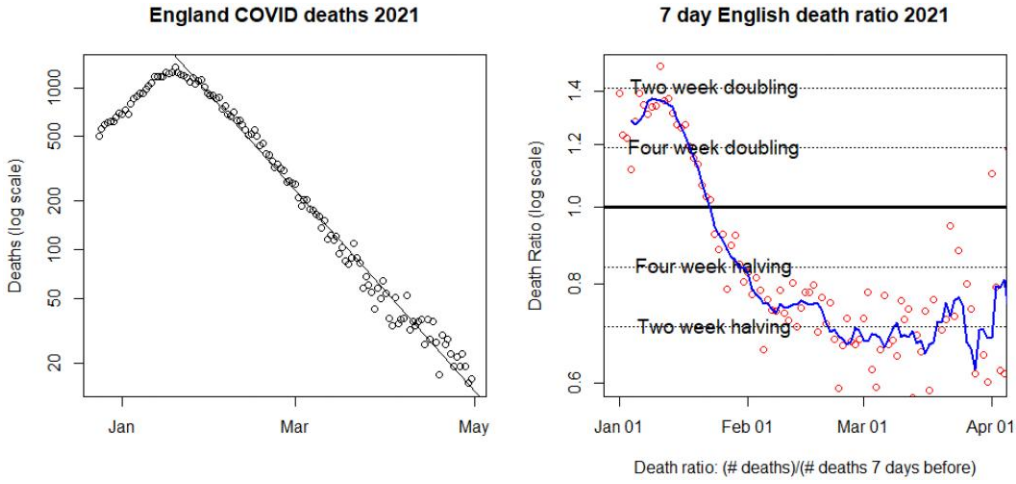



**Figure 5.** England death data January–April 2021 by date of occurrence. a) Absolute numbers of deaths (log scale), line fitted to period of exponential decay using `lm()` in R. b) Daily death ratio (deaths on a given day divided by deaths 7 days before, to smooth out weekly effects). Data from archived UKHSA COVID dashboard (UK Health Security Agency (UKHSA), 2025).

changes in infection numbers (see Figure 5 for death data), suggesting that estimates of trends in growth periods were broadly reliable.

### 1.2 Reporting lag

Further, while more precise, the ONS survey suffered from reporting lag. For example, on 11th December 2020 (Office for National Statistics (ONS), 2020), it reported an overall reduction in prevalence in England and only a modest rise in London (Figure 3). These delayed estimates, referring to 29th November to 5th December, held for the next week and fed into the official consensus 11th–17th December R number estimates UK Health Security Agency (UKHSA) (2022a) of 0.8–1.0 (England) and 0.9–1.1 (London). However, contemporaneous analysis of cases, which are closer to the key incidence metric than to PCR prevalence, gave more timely warning of the alpha wave (Figure 4; Johnson, 2020a, 2020b).

## 2 Section 5: Lockdowns

While it's interesting to consider the possibility that action just short of a lockdown could cause infections to peak ($R$ to go under 1), a key statement (Wood et al., 2025, P 20) is 'lockdowns

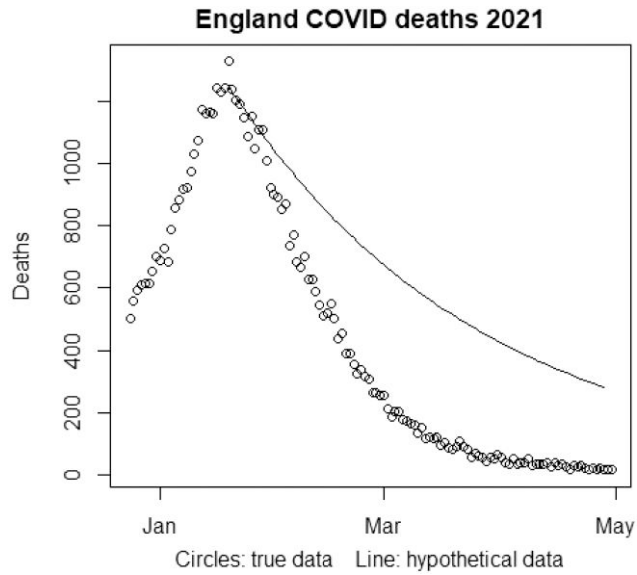



**Figure 6.** True death data (UK Health Security Agency (UKHSA), 2025) plotted as circles (linear scale, estimated 28% weekly decline), along with hypothetical trajectory plotted as a line (modelled 10% weekly decline).

having further suppressed infections, causing infection waves to subside more quickly than might have otherwise occurred'.

To illustrate this, Figure 5 shows that under lockdown deaths in early 2021 were well modelled by a consistent exponential fall (straight line on log scale), with the fitted line representing a 28% weekly decline in deaths and greater later variability due to random fluctuations.

As a result, while deaths fell from a peak of 1,328 on 19th January, there were a total of 33,773 deaths in the next 100 days. Obviously, we cannot know what would have happened without lockdown, but presumably greater social mixing would have led to a slower rate of exponential decay. For illustration, even accepting the idea that $R < 1$ might be possible without lockdown, Figure 6 shows a hypothetical trajectory of a slower (10% weekly) decline resulting in 65,100 deaths over the same period.

Applying the NICE criteria of Wood et al. (2025, Section 1), it would be reasonable to spend £9bn to avert these extra ~ 30,000 deaths, with presumably some significant additional monetary value from reduced hospitalizations (including greater available hospital capacity to treat other conditions). While this figure may be lower than the average £ 3.7bn/month spent on the furlough scheme alone House of Commons Library (2021), it's not hugely different (and obviously different hypothetical rates of decline would lead to different figures).

Further, it illustrates a general principle that the benefits of lockdown are likely to decay exponentially over time, while costs accumulate linearly (Wood et al., 2025, Figure 2). Indeed, 32,052 of these 33,773 deaths occurred in the first 60 days. This crude analysis might perhaps suggest that the first 2 months of 2021 lockdown met the NICE criteria cited by Wood et al. (2025), but the benefit of later measures was less clear.

In general, such analysis might imply that the optimal strategy may involve short and intense periods of maximal suppression—giving weight to the conclusion of Wood et al. (2025, P. 20) that 'the decision to continue suppression well into the summer of 2020 does not seem optimal'. However, unlike Wood et al. (2025, Figure 1), it feels that any analysis of deaths and consequences should take into account that very high numbers of COVID deaths were concentrated into relatively short periods of time.

*Conflict of interests:* None declared.

## References

BBC News. (2020). Coronavirus: 'widespread virus growth across the country'. Retrieved March 31, 2025, from https://www.bbc.co.uk/news/health-54206705.

House of Commons Library. (2021). Coronavirus job retention scheme: statistics. Retrieved March 31, 2025, from https://commonslibrary.parliament.uk/research-briefings/cbp-9152/.
Johnson O. (2020a). Tweet. 9th December. Retrieved March 31, 2025, from https://twitter.com/BristOliver/status/1336704276878516226.
Johnson O. (2020b). Tweet. 14th December. Retrieved March 31, 2025, from https://twitter.com/BristOliver/status/1338450287040876545.
Office for National Statistics (ONS). (2020). Coronavirus (COVID-19) Infection Survey, UK. Retrieved March 31, 2025, from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/11december2020.
UK Health Security Agency (UKHSA). (2022a). The R value and growth rate. Retrieved March 31, 2025, from https://www.gov.uk/guidance/the-r-value-and-growth-rate.
UK Health Security Agency (UKHSA). (2022b). Weekly statistics for NHS Test and Trace (England): 2 to 15 June 2022, 15th June 2022. Retrieved March 31, 2025, from https://www.gov.uk/government/publications/weekly-statistics-for-nhs-test-and-trace-england-2-to-15-june-2022.
UK Health Security Agency (UKHSA). (2025). COVID-19 Archive data download. Retrieved March 31, 2025, from https://ukhsa-dashboard.data.gov.uk/covid-19-archive-data-download.
Wood S. N., Wit E. C., McKeigue P. M., Hu D., Flood B., Corcoran L., & Abou Jawad T. (2025). Some statistical aspects of the Covid-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society (to Appear)*.

https://doi.org/10.1093/jrsssa/qnaf094
Advance access publication 28 June 2025



# Victor Richmond Jose's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Victor Richmond R. Jose**

McDonough School of Business, Georgetown University, Washington, DC, USA

*Address for correspondence*: Victor Richmond R. Jose, Operations and Analytics Area, McDonough School of Business, Georgetown University, 3700 O St NW, Washington, DC 20057, USA. Email: vrj2@georgetown.edu

I congratulate the authors for an important retrospective that provides important guidance for statisticians and policy-makers responding to future pandemics.

One of the fundamental challenges in presenting to a general audience is the heterogeneity in the public's ability to properly interpret risk (Gordon-Lubitz, 2003). One way to perhaps improve future dissemination of risk is the use of more probabilistic approaches (e.g. distribution-based, interval-oriented, margin of errors, etc.). Though this requires a greater degree of numeracy for full comprehension, it provides a clearer picture to those in the scientific community. It also can be used to distill simpler information or cases for summarizing, dashboarding, or perhaps emotional-based messaging (e.g. by focusing on extreme outcomes) when addressing other audiences. Aside from the approach, it may be of interest to study which data to present since risk perceptions based on objective measure of situational severity (e.g. number of confirmed COVID-19 cases) can significantly differ within subgroups of the population (Schneider et al., 2021). This may also require different forms of statistical analysis and risk communication based on the audience.

Second, I wanted to comment on the authors' observation about the over-reliance and disproportionate weight of support placed on models. I believe that two additional factors are connected to this—the illusions of objectivity and persuasiveness. As the authors note, the online environment may not be the ideal environment to propose novel and untested theories and models. This led to the heavier use of existing models that reduced the burden among scientists/statisticians

for 'proof of concept.' It also made the choice appear less arbitrary since such models have already been studied. However, it can be difficult to validate these models since wrong models often appear to be more persuasive when there is a lack of data or too much noise in the data (Schwartzstein & Sunderam, 2021).

Lastly, I wondered what the authors think were *some* of the things (related to statistical aspects of the COVID-19 response) the UK did well compared with other countries. These could equally be as informative for future pandemic response as the questionable approaches highlighted in these discussions.

I thank the authors for making this important contribution by documenting this unique period of history and continuing this important discussion on how statistical practice can contribute to healthcare policy decision-making. I hope this discussion continues, especially related to how statistical concepts can be properly and effectively communicated to the public.

*Conflicts of interest:* The author declares that he has no conflicts of interest related to this discussion.

## References

Gordon-Lubitz, R. (2003). Risk communication: Problems of presentation and understanding. *JAMA*, *289*(1), 95. https://doi.org/10.1001/jama.289.1.95

Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., Spiegelhalter, D., & van der Linden, S. (2021). COVID-19 risk perception: A longitudinal analysis of its predictors and associations with health protective behaviours in the United Kingdom. *Journal of Risk Research*, *24*(3–4), 294–313. https://doi.org/10.1080/13669877.2021.1890637

Schwartzstein, J., & Sunderam, A. (2021). Using models to persuade. *American Economic Review*, *111*(1), 276–323. https://doi.org/10.1257/aer.20191074

https://doi.org/10.1093/jrsssa/qnaf086
Advance access publication 27 June 2025

# Matt Keeling's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Matt J. Keeling**

SBIDER, Mathematics Institute and School of Life Sciences, University of Warwick, Coventry, UK

*Address for correspondence:* Matt J. Keeling, SBIDER, Mathematics Institute and School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK. Email: m.j.keeling@warwick.ac.uk

Five years after the COVID-19 pandemic emerged, Wood et al. have pulled together an eclectic mix of comments and criticisms of the activities undertaken during 2020–21. Those who were directly involved with providing advice during the pandemic, would be the first to agree that the statistics produced (generally to meet extremely tight policy requirements) were often imperfect, and greater practical assistance from the UK statistical community would have been of benefit.

A diversity of subjects is covered, but I will restrict my attention to Wood's comments on the modelling of the Omicron wave (Figure 15) where the work by Warwick University is mentioned. There are multiple instances when modelling was highly successful (circuit-breakers lockdowns (Keeling et al. 2021), vaccination prioritization (Moore et al. 2021), or short-term forecasting (Manley et al. 2024)), yet Wood et al. pick a time and metric where the models perform badly.

(Due to delays in the system, the Warwick modelling in the 19th December SPI-MO report used case data up to 10th December—just two weeks after Omicron was first detected in the UK.) Wood's comparison comes with the considerable benefit of hindsight, and does not consider either how the statistical fit of such models could be improved, or the difficulties of fitting a model to very early outbreak data (Keeling & Dyson (2024)). The authors' state that the Omicron wave should provide an ideal test as there were no interventions, however, this ignores the government's 'Plan-B', the numerous calls in the media for greater caution and greater testing and the disruption to normal mixing patterns caused by the Christmas holidays. The prediction of the Omicron wave was also complicated by the far shorter generation time of Omicron compared with previous variants and the lower severity of illness (historically we had experienced increasing severity with the Alpha and Delta variants). As highlighted in Keeling and Dyson (2024), we believe accurately assessing the scale of the Omicron wave was infeasible from early data due to these differences between Omicron and previous variants, and the speed of invasion.

In summary, this paper is a missed opportunity. It does not provide the epidemiological modelling community with additional statistical tools or insights; it does not offer a series of lessons to be learned from the COVID-19 pandemic; nor does it suggest best-practices when dealing with future outbreaks. I hope that when the next epidemic occurs, the academic community will rise to the statistical challenges, and provide much needed support to public health advisors.

*Conflict of interest:* none declared.

## References

Keeling, M. J., & Dyson, L. (2024). A retrospective assessment of forecasting the peak of the SARS-CoV-2 omicron BA.1 wave in England. *PLoS Computational Biology*, *20*(9), e1012452. https://doi.org/10.1371/journal.pcbi.1012452

Keeling, M. J., Guyver-Fletcher, G., Dyson, L., Tildesley, M. J., Hill, E. M., Medley, G. F. (2021). Precautionary breaks: Planned, limited duration circuit breaks to control the prevalence of SARS-CoV2 and the burden of COVID-19 disease. *Epidemics*, *37*, 100526. https://doi.org/10.1016/j.epidem.2021.100526

Manley, H., Bayley, T., Danelian, G., Burton, L., Finnie, T., Charlett, A., Watkins, N. A., Birrell, P., De Angelis, D., Keeling, M., Funk, S., Medley, G., Pellis, L., Baguelin, M., Ackland, G. J., Hutchinson, J., Riley, S., Panovska-Griffiths, J. (2024). Combining models to generate consensus medium-term projections of hospital admissions, occupancy and deaths relating to COVID-19 in England. *Royal Society Open Science*, *11*(5), 231832. https://doi.org/10.1098/rsos.231832

Moore, S., Hill, E. M., Dyson, L., Tildesley, M. J., Keeling, M. J. (2021). Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. *PLoS Computational Biology*, *17*(5), e1008849. https://doi.org/10.1371/journal.pcbi.1008849

https://doi.org/10.1093/jrsssa/qnaf073
Advance access publication 26 June 2025

# Kuldeep Kumar's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood *et al.*

**Kuldeep Kumar**

Centre for Data Analytics, Bond University, Gold Coast, Queensland 4229, Australia

*Address for correspondence*: Kuldeep Kumar, Centre for Data Analytics, Bond University, Gold Coast, Queensland 4229, Australia. Email: kkumar@bond.edu.au

I would like to congratulate Simon Wood and his team for bringing out such a comprehensive paper on the UK's Covid-19 response highlighting major statistical oversights in modelling, data presentation, and media reporting, emphasizing the need for more balanced and transparent approaches in future public health crises. One of the major concerns which has been overlooked in this paper is 'dark data'. It may refer to data that were missing, inaccessible, not collected, or not used effectively in public health decision-making. Many Covid-19 cases, especially asymptomatic or mild ones, were never tested or recorded—particularly early in the pandemic. Consideration of dark data is very important in risk assessment, model prediction and policy effectiveness evaluations. My other concern is authors have used a lifetable-based method for calculating excess death, but have they considered the reduction in mortality because of fewer accidents and other causes because of the lockdowns? Finally, I would like to draw the attention of the authors to a recent paper by Shalabh *et al.* (2024) which uses data driven statistical models for modelling COVID-19 data. This approach is particularly beneficial in the case when the exact distribution of the data is unknown, or parameters involved in the probability density function (if we assume a certain form of the distribution) are unknown.

*Conflicts of interest:* None declared.

## Reference

Shalabh, Dhar, S. S., & Rajeshbhai, S. P. (2024). Statistical data-driven modelling and forecasting: An application to COVID-19 pandemic. *Annals of Data Science*, 1–24. https://doi.org/10.1007/s40745-024-00583-8

https://doi.org/10.1093/jrsssa/qnaf084
Advance access publication 27 June 2025

# Yuhang Liu, Weihao Wang, Weng-Kee Wong, and Wei Zhu's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Yuhang Liu[1], Weihao Wang[1], Wei Zhu[1] and Weng Kee Wong[2]**

[1]Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, NY, USA
[2]Department of Biostatistics, University of California at Los Angeles, Los Angeles, CA, USA

*Address for correspondence*: Weng Kee Wong, Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, California 90095-1772, USA. Email: wkwong@ucla.edu

The modelling of non-pharmaceutical interventions (NPIs) during the COVID-19 pandemic has prompted many thought-provoking discussions surrounding methodology, interpretation, and the implications for rapid policy-making under uncertainty. Wood et al. (2025) provide a thoughtful critique that brings valuable attention to potential biases in estimating intervention effects, particularly the concerns that early interventions, such as lockdowns, may appear disproportionately effective simply due to their position in the intervention timeline. This highlights the need to scrutinize underlying model assumptions, especially when such analyses inform high-stakes societal decisions.

Wood et al., referencing the influential work of Flaxman et al. (2020), point out that in many European countries, lockdowns were among the earliest interventions imposed. This

chronological positioning could lead models to overstate their impact. However, Liu et al. (2024) applied the same modelling framework to data from the U.S., analysing intervention effects across multiple states. They found that stay-at-home orders consistently ranked among the most effective NPIs—even when implemented later than other measures such as business closures or restrictions on gatherings. These findings suggest that the perceived effectiveness of lockdown-type measures cannot be solely attributed to their early implementation. This raises a possible counterpoint: could there be intrinsic features of lockdown-type measures that contribute to their consistent effectiveness, beyond timing alone? If so, how might future models better disentangle timing from intrinsic intervention efficacy?

Another key point of debate concerns the assumption that the effective reproduction number (Rt) changes in a stepwise rather than continuous fashion. While this simplification may not fully capture the complexity of real-world transmission dynamics, it represents a conscious trade-off aimed at enhancing policy relevance. The primary objective is to identify statistically significant shifts in transmission that coincide with specific interventions, rather than to reconstruct the epidemic curve with granular precision. Nonetheless, we acknowledge the potential of such assumptions to obscure important dynamics in settings with gradual or staggered changes.

These methodological considerations reflect a broader tension in epidemic modelling: the trade-off between realism and actionable insight. Both stepwise and continuous modelling approaches have their merits and limitations, depending on the specific goals of the analysis. Wood's critique appropriately calls for greater sensitivity testing and more cautious interpretation of results. Moving forward, model refinement should aim to preserve analytical rigour while enhancing usability for public health decision-makers operating under uncertainty.

*Conflicts of interest:* none declared.

## References

Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M.; Imperial College COVID-19 Response Team, Ghani, A. C., Donnelly, C. A., Riley, S., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Liu, Y., Wang, W., Wong, W. K., & Zhu, W. (2024). Effectiveness of non-pharmaceutical interventions for COVID-19 in USA. *Scientific Reports*, *14*(1), 21387. https://doi.org/10.1038/s41598-024-71984-1

Wood, S. N., Wit, E. C., McKeigue, P. M., Hu, D., Flood, B., Corcoran, L., & Jawad, T. A. (2026). Some statistical aspects of the Covid-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *189*(1), 1–31. https://doi.org/10.1093/jrsssa/qnaf092

https://doi.org/10.1093/jrsssa/qnaf100
Advance access publication 4 July 2025



# Antonello Maruotti, Saleh Ibrahim, Emilio Porcu, and Horst Simon's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.'

**Antonello Maruotti[1], Emilio Porcu[2,3], Saleh Ibrahim[3,4] and Horst Simon[3]**

[1]Department GEPLI, LUMSA University, Rome, Italy
[2]Department of Mathematics, Khalifa University, Abu Dhabi, United Arab Emirates

[3]ADIA Lab, ADGM Tower, Abu Dhabi, United Arab Emirates
[4]College of Health and Life Sciences, Khalifa University, Abu Dhabi, United Arab Emirates
*Address for correspondence*: Emilio Porcu, Department of Mathematics, Khalifa University, Abu Dhabi 20000, United Arab Emirates. Email: georgepolya01@gmail.com

We congratulate the authors for their thoughtful analysis of the statistical challenges in shaping policy decisions during the COVID-19 pandemic, especially regarding monitoring and risk assessment. Their work highlights how data analysis must address not just *how* models are used, but *why*—focusing on inference, intervention evaluation, and communication with both stakeholders and the public.

A critical issue remains the quality and suitability of data and models. Under the 'garbage in, garbage out' paradigm, poor data inputs, unrealistic assumptions, or oversimplified epidemiological features can heavily lead to unreliable results (Ioannidis et al., 2020). While compartmental models are foundational in epidemiology, their reliability depends on high-quality, up-to-date data. Data-driven approaches (Mingione et al., 2024) may offer more robust alternatives, especially under uncertainty or limited data.

This concern is well illustrated by reproduction number, $R_t$, estimation in Italy. The serial interval, a fundamental input in the estimation process of the reproduction number, was derived from just 90 case pairs in Lombardy during February 2020 (Cereda et al., 2021), using a fixed Gamma distribution that remained unchanged across time and regions. This one-size-fits-all assumption introduced substantial regional variability and questionable reliability in $R_t$ estimates (Maruotti et al., 2021).

Cross-country comparisons pose further challenges. While excess mortality is a common metric, it requires model-based analysis (Alaimo Di Loro et al., 2025; Maruotti et al., 2022; Scrucca, 2022; Verbeeck et al., 2023), rather than simple historical averages comparisons. Even then, international comparisons can be biased by differing socio-demographic structures. Standardized mortality rates and their evolution over time may offer a more accurate indicator for such comparisons.

The evaluation of non-pharmaceutical interventions (NPIs), especially lockdowns, must consider timing, context, and which outcomes are being measured. For instance, some NPIs may affect certain epidemiological indicators but not others, with impacts that are often short-lived or highly dependent on the prevailing epidemiological context (Pelagatti & Maranzano, 2021), yet they may still carry significant social or economic consequences. These trade-offs merit deeper, context-specific investigation.

Finally, it's important to remember that a good model fit does not guarantee predictive accuracy. The influence of individual statistical choices—whether in data selection, model inputs, or assumptions—carries weighty implications, particularly when these estimates guide policymaking.

Two important questions remain open: Which countries managed the pandemic most effectively? And what statistical best practices should be adopted to better prepare for future public health crises?

*Conflicts of interest:* None declared.

## References

Alaimo Di Loro P., Böhning D., & Sahu S. K. (2025). A Bayesian spatio-temporal Poisson auto-regressive model for the disease infection rate: Application to COVID-19 cases in England. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *74*(3), 551–575.

Cereda D., Manica M., Tirani M., Rovida F., Demicheli V., Ajelli M., Poletti P., Trentini F., Guzzetta G., Marziano V., Piccarreta R., Barone A., Magoni M., Deandrea S., Diurno G., Lombardo M., Faccini M., Pan A., Bruno R., … Merler S. (2021). The early phase of the Covid-19 epidemic in Lombardy, Italy. *Epidemics*, *37*(3), 100528. https://doi.org/10.1016/j.epidem.2021.100528

Ioannidis J. P. A., Cripps S., & Tanner M. A. (2020). Forecasting for COVID-19 has failed. *International Journal of Forecasting*, *38*, 423–438. https://doi.org/10.1016/j.ijforecast.2020.08.004

Maruotti A., Ciccozzi M., & Divino F. (2021). On the misuse of the reproduction number in the Covid-19 surveillance system in Italy. *Journal of Medical Virology*, *93*(5), 2569. https://doi.org/10.1002/jmv.v93.5

Maruotti A., Jona-Lasinio G., Divino F., Lovison G., Ciccozzi M., & Farcomeni A. (2022). Estimating Covid-19-induced excess mortality in Lombardy, Italy. *Aging Clinical and Experimental Research*, *34*(2), 475–479. https://doi.org/10.1007/s40520-021-02060-1

Mingione M., Alaimo Di Loro P., & Maruotti A. (2024). A useful parametric specification to model epidemiological data: Revival of the Richards' curve. *Statistical Methods in Medical Research*, *33*(8), 1473–1494. PMID: 39360928. https://doi.org/10.1177/09622802241262522

Pelagatti M., & Maranzano P. (2021). Assessing the effectiveness of the Italian risk-zones policy during the second wave of Covid-19. *Health Policy*, *125*(9), 1188–1199. https://doi.org/10.1016/j.healthpol.2021.07.011

Scrucca L. (2022). A covindex based on a gam beta regression model with an application to the Covid-19 pandemic in Italy. *Statistical Methods & Applications*, *31*(4), 881–900. https://doi.org/10.1007/s10260-021-00617-y

Verbeeck J., Faes C., Neyens T., Hens N., Verbeke G., Deboosere P., & Molenberghs G. (2023). A linear mixed model to estimate Covid-19-induced excess mortality. *Biometrics*, *79*(1), 417–425. https://doi.org/10.1111/biom.v79.1

https://doi.org/10.1093/jrsssa/qnaf096
Advance access publication 7 July 2025

# Stuart McDonald's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Stuart McDonald**

Lane Clark & Peacock LLP, Actuarial Department, London, UK

*Address for correspondence*: Stuart McDonald, Lane Clark & Peacock LLP, Actuarial Department, 95 Wigmore Street, London, W1U 1DQ, UK. Email: stuart.mcdonald@lcp.uk.com

It can be valuable, with the benefit of hindsight, to analyse the way that statistics were presented and used in a crisis such as the Covid-19 pandemic, so as to learn lessons for the future. However, there are several points I disagree with in this paper (Wood et al. 2025). In the interests of brevity, I will focus my comments on the analysis of excess deaths.

The authors are right to highlight the importance of allowing for changes to the age structure and size of the population when calculating excess deaths. The iterated life table approach described in the paper has much in common with the approach taken by the Continuous Mortality Investigation (CMI) in their weekly Mortality Monitor publications (Continuous Mortality Investigation).

The CMI estimates 151,700 excess deaths (Continuous Mortality Investigation, 2024) from the onset of the pandemic until the end of 2022. This is 60% higher than the authors' estimate of 'around 95 thousand total excess deaths'. This significant discrepancy (56,700 deaths) arises from two key differences.

Firstly, and most materially, the authors calculate the number of expected deaths based on mortality rates in 2017–19, with no adjustment for subsequent improvements in mortality. Although the rate of improvement has been slower since 2011 compared to previous decades, mortality rates continued to fall up until the onset of the pandemic (BBC News, 2025), and it is reasonable to assume that they would have continued to fall thereafter.

Actual mortality rates were typically above the authors' baseline in 2017 and early 2018, but below the baseline thereafter until the onset of the pandemic. Close examination of the authors' Figure 4 bears this out. Figure 1 seeks to make this point more clearly, showing that deviations between observed mortality rates and the authors' baseline are not random. Mortality rates were significantly lower than the baseline in the 2 years before April 2020.

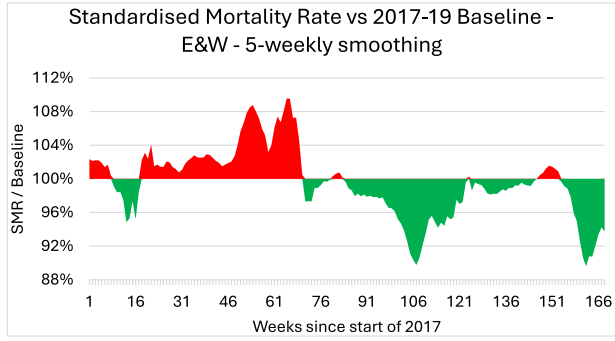



**Figure 1.** Five-week rolling average mortality rate for England & Wales from the start of 2017 until Week 12 of 2020, relative to a 2017–19 baseline.

The second point is related to the first. The authors calculate excess deaths from week 1 of 2020, rather than from the point when Covid-19 led to an increase in UK death rates. Because their baseline is too high, they record a significant 'negative excess' in the first 3 months of 2020, i.e. their measure of excess starts from a deficit of nearly 10,000 deaths at the onset of the pandemic. This ought to have been a red flag as to the appropriateness of the baseline.

For both these reasons, I prefer the CMI analysis to the analysis in this article, and the CMI method might have been a more appropriate comparator than a superseded Office for National Statistics (ONS) method. The authors criticism of the ONS approach is somewhat redundant as ONS made the same points when introducing their updated method (Office for National Statistics (ONS), 2024). In any case, the authors have used the wrong ONS numbers, the 167,356 number they cite is for England and Wales, not for the UK (Office for National Statistics (ONS), 2023).

*Conflicts of interest:* S.M. is a Partner at Lane Clark & Peacock LLP, a Fellow of the Institute and Faculty of Actuaries and Deputy Chair of the Continuous Mortality Investigation.

## Data Availability

The data underlying Figure 1 will be shared on reasonable request to the corresponding author.

## References

BBC News. (2025). *UK death rate 'reaches record low'*. (3 March 2025). https://www.bbc.co.uk/news/articles/c8j09vn2x2ko.

Continuous Mortality Investigation. *Mortality Monitor*. (26 March 2025). https://www.actuaries.org.uk/learn-and-develop/continuous-mortality-investigation/other-cmi-outputs/mortality-monitor.

Continuous Mortality Investigation. (2024). *Mortality Monitor*. Week 1 of 2024. https://www.actuaries.org.uk/documents/england-wales-mortality-monitor-covid-19-update-week-1-2024.

Office for National Statistics (ONS). (2023). 9 March 2023. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/articles/excessdeathsinenglandandwalesmarch2020todecember2022/2023-03-09.

Office for National Statistics (ONS). (2024). 20 February 2024. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/estimatingexcessdeathsintheukmethodologychanges/february2024.

Wood S. N., Wit E. C., McKeigue P. M., Hu D., Flood B., Corcoran L., & Abou Jawad T. (2026). Some statistical aspects of the Covid-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *189*(1), 1–31. https://doi.org/10.1093/jrsssa/qnaf092

https://doi.org/10.1093/jrsssa/qnaf091
Advance access publication 28 June 2025

# Geert Molenberghs' contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Geert Molenberghs**[1,2]

[1]I-BioStat, Data Science Institute, Universiteit Hasselt, BE-3500 Hasselt, Belgium
[2]I-BioStat, L-BioStat, KU Leuven, BE-3000 Leuven, Belgium

*Address for correspondence*: Geert Molenberghs, I-BioStat, Universiteit Hasselt and KU Leuven, BE-3500 Hasselt, Belgium. Email: geert.molenberghs@uhasselt.be



We have been active on various government advisory and other policy boards—predominantly in Belgium but also at European level—as well as in pandemic scientific research (Molenberghs, 2022; Molenberghs et al., 2020). It is not our experience that risks were exaggerated to obtain compliance; rather there were periods where it was hard to convince policymakers that taking measures was necessary, for example, when a new peak was emerging, because measures early taken can be less stringent and shorter in duration. It surely is important to distinguish between individual and collective risks, with measures taken in (age) strata with less risk also in favour of more vulnerable and older population members. Early implementation of measures is simultaneously in the best interest of public health, economy, and well-being (Brackx et al., 2023; Loedy et al., 2023), rather than favouring one of the other. For this reason, policy boards were multi-disciplinary, in spite of the common but incorrect perception that they consisted solely of health and quantitative experts. Accompanying economic support programs are required, which happened in Belgium. Many mathematical models used did take random person-to-person variability into account, as well as other features of the infection, including pre-symptomatic transmission, evolving properties due to variants, and vaccination effects (Abrams et al., 2021). Scientists continually communicated that the models are useful for scenario analysis as one component for policy decisions, which differs from prediction due to the possibility to intervene and hence change the course of the pandemic. Belgium carefully reported COVID-19 mortality (regardless of venue, test, and clinically confirmed) and for 2020 amounted to 19,837 (against the background of roughly 110,000 yearly deaths); excess mortality for 2020 was estimated to be 17,177, with much lower figures for 2021 and 2022, and a further return to normality in subsequent years (Adelwin Natalia et al., 2024; Molenberghs et al., 2022; Verbeeck et al., 2023), adding nuance to the population-aging argument (for reference: 108,745 in 2019; 111,255 in 2023). It is true that life expectancy dropped sharply in 2020 but rebounded in years following; these quantities should be jointly monitored. Many measures were taken to minimize the impact of measures and towards taking targeted measures rather than general lockdowns (Brackx et al., 2022; Stijven et al., 2023; Verbeeck et al., 2021). Of course, it is important to draw proper lessons from the pandemic (Angeli et al., 2025; Fajgenblat et al., 2024; Grieve et al., 2023).

*Conflicts of interest:* None declared.

## References

Abrams S., Wambua J., Santermans E., Willem L., Kuylen E., Coletti P., Libin P., Faes C., Petrof O., Herzog S. A., Beutels P., & Hens N. (2021). Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. *Epidemics*, *35*(163), 100449. https://doi.org/10.1016/j.epidem.2021.100449

Adelwin Natalia Y., Verbeeck J., Faes C., Neyens T., & Molenberghs G. (2024). Unraveling the impact of the COVID-19 pandemic on the mortality trends in Belgium between 2020–2022. *BMC Public Health*, *24*(1), 2916. https://doi.org/10.1186/s12889-024-20415-x

Angeli L., Pereira Caetano C., Franco N., Coletti P., Faes C., Molenberghs G., Beutels P., Abrams S., Willem L., & Hens N. (2025). Insights on the role of children in COVID-19 pandemic in Belgium: A longitudinal sensitivity analysis. *Nature Communications*, *16*(1), 2230. https://doi.org/10.1038/s41467-025-57087-z

Brackx F., Molenberghs G., & De Smedt B. (2023). A dashboard for the evaluation of the effect of school closures on wellbeing. *Archives of Public Health*, *81*(1), 178. https://doi.org/10.1186/s13690-023-01114-2

Brackx F., Vanongeval F., Natalia Y. A., Molenberghs G., & Steenbergen T. (2022). The effect of transborder mobility on COVID-19 incidences in Belgium. *International Journal of Environmental Research and Public Health*, *19*(16), 9968. https://doi.org/10.3390/ijerph19169968

Fajgenblat M., Molenberghs G., Verbeeck J., Willem L., Crèvecoeur J., Faes C., Hens N., Deboosere P., Verbeke G., & Neyens T. (2024). Evaluating the direct effect of vaccination and non-pharmaceutical interventions during the COVID-19 pandemic in Europe. *Communications Medicine*, *4*(1), 178. https://doi.org/10.1038/s43856-024-00600-0

Grieve R., Yang Y., Abbott S., Babu G., Bhattacharyya M., Dean N., Evans S., Jewell N., Langan S. M., Lee W., Molenberghs G., Smeeth L., Williamson E., & Mukherjee B. (2023). The importance of investing in data, models, experiments, team science and public trust to help policymakers prepare for the next pandemic. *PLOS Global Public Health*, *3*(11), e0002601. https://doi.org/10.1371/journal.pgph.0002601

Loedy N., Coletti P., Wambua J., Hermans L., Willem L., Jarvis C. I., Wong K. L. M., Edmunds W. J., Robert A., Leclerc Q. J., Gimma A., Molenberghs G., Beutels P., Faes C., & Hens N. (2023). Longitudinal social contact data analysis: Insights from 2 years of data collection in Belgium during the COVID-19 pandemic. *BMC Public Health*, *23*(1), 1298. https://doi.org/10.1186/s12889-023-16193-7

Molenberghs G. (2022). Biostatistics and the COVID-19 pandemic in Belgium, in 2020 and 2021. *Statistique et Société*, *10*, 31–44. https://doi.org/10.4000/statsoc.482

Molenberghs G., Buyse M., Abrams S., Hens N., Beutels P., Faes C., Verbeke G., Van Damme P., Goossens H., Neyens T., Sereina Herzog T. H., Pepermans K., Alonso Abad A., Van Keilegom I., Speybroek N., Legrand C., De Buyser S., & Hulstaert F. (2020). Infectious diseases epidemiology, quantitative methodology, and clinical research in the midst of the COVID-19 pandemic: Perspective from a European country. *Controlled Clinical Trials*, *99*, 106189. https://doi.org/10.1016/j.cct.2020.106189

Molenberghs G., Faes C., Verbeeck J., Deboosere P., Abrams S., Willem L., Aerts J., Theeten H., Devleesschauwer B., Bustos Sierra N., Renard F., Herzog S., Lusyne P., Van der Heyden J., Van Oyen H., Van Damme P., & Hens N. (2022). Belgian COVID-19 mortality, excess deaths, number of deaths per million, and infection fatality rates (9 March–28 June 2020). *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin*, *27*(7), 7. https://doi.org/10.2807/1560-7917.ES.2022.27.7.2002060

Stijven F., Verbeeck J., & Molenberghs G. (2023). Comparing COVID-19 incidences longitudinally per economic sector against the background of preventive measures and vaccination. *Biometrics*, *79*(3), 2516–2524. https://doi.org/10.1111/biom.13766

Verbeeck J., Faes C., Neyens T., Hens N., Verbeke G., Deboosere P., & Molenberghs G. (2023). A linear mixed model to estimate COVID-19-induced excess mortality. *Biometrics*, *79*(1), 417–425. https://doi.org/10.1111/biom.v79.1

Verbeeck J., Vandersmissen G., Peeters J., Klamer S., Hancart S., Lernout T., Dewatripont M., Godderis L., & Molenberghs G. (2021). Confirmed COVID-19 cases per economic activity during Autumn wave in Belgium. *International Journal of Environmental Research and Public Health*, *18*(23), 12489. https://doi.org/10.3390/ijerph182312489

https://doi.org/10.1093/jrsssa/qnaf093
Advance access publication 1 July 2025



# Professor Lorenzo Pellis' contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Lorenzo Pellis**

Department of Mathematics, The University of Manchester, Manchester, UK

*Address for correspondence*: Lorenzo Pellis, Department of Mathematics, The University of Manchester, Alan Turing Building, Oxford Road, Manchester M13 9PL, UK. Email: lorenzo.pellis@manchester.ac.uk

For brevity, I will focus only on the lockdown discussion. Beyond a generalized criticism, I am still unclear what the authors specifically disagree with. Would have they not entered Lockdown 1? I do not think so (from paragraph 2 in Discussion), but if this were the case, I would point them to Pellis et al. (2021) or (https://www.gov.uk/government/publications/university-of-manchester-uk-covid-19-predictions-23-march-2020). Conversely, I might agree that better learning from Lockdown 1 could have helped better decisions about following lockdowns, though this is still not straightforward (see opinions below). Ultimately, however, all analyses presented are retrospective, with the luxury of data and hindsight: how can this work be used for real-time response in the next pandemic?

Scientific considerations:

- Against what time is R plotted in Figure 10? The threshold of 1 is crossed roughly at lockdown, but the incidence peaks before then.
- I agree assuming constant transmission until lockdown attributes all impact of control on lockdown *by design*, which the GAM + deconvolution approach does not. However, GAMs are not assumption-free: they just rely on different assumptions, like using splines or the number of knots (incidentally, not mentioned in the paper). In particular, the reliance on splines means that the most likely reconstructed incidence fitting some smoothed delayed data is *by design* sufficiently smooth, even when synthetic data come from a sudden change in transmission (Figure 1).
- The authors praise their approach for displaying a rebound in R after lockdown in Figure 14, which they suggest reflects the mechanistic argument in Section 4.2. However, this is incorrect, the GAM being mechanism-agnostic: instead, I suspect this might be due to the death data decline flattening out in May in Figure 10. Furthermore, our group used GAMs extensively during the pandemic, and despite their usefulness I would be hesitant to draw strong conclusions from this rebound, as GAMs tend to over-display oscillations purely due to the nature of splines (Figure 2).
- Finally, fitting a 'static' dataset is very different from real-time estimation, e.g. in Figure 1 the incidence curve 'turns around' before the sudden change in trend to be as close as possible to all data points—i.e. information about 'now' is borrowed from future data points. Figure 2 shows how estimates of R change retrospectively as more points after a sudden change in trend 'become available'.

Further personal opinions:

- Insights might not easily transfer between lockdowns, as conditions change (new variants, unusual interaction patterns around Christmas, etc.).
- Fine control of fast-growth systems with long delays is *really* hard—a *substantial* margin for manoeuvre is needed to prevent losing control.
- The tier system (not informed by any SPI-M analysis) could have worked better if areas were moved to a higher tier not based on prevalence, but on trend (i.e. seeing cases stably growing, when low prevalence still offers room for manoeuvre). As implemented, the system merely

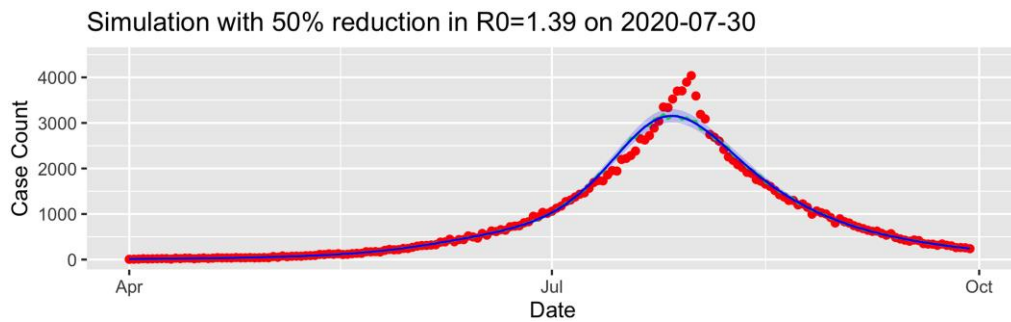



**Figure 1.** Simulation from an SEIR model with a drop in R from 1.39 to 0.7 on 31 July, adapted from Figure 2 in Pang et al. (2025) to a population is 60 million, so no depletion of susceptibles is perceived. Data (dots) are daily numbers entering state I, fitted by a GAM (continuous line, with 95% CrI shown as lighter band).

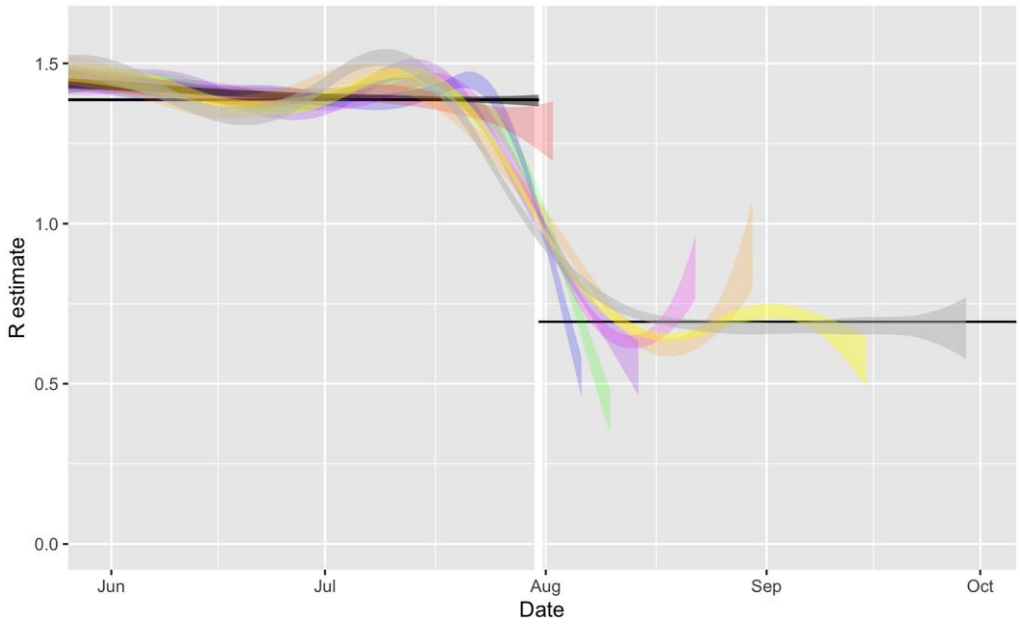



**Figure 2.** Estimate of R from the GAM fit in Figure 1 (zoomed x-scale), as in Figure 4 (bottom-right) of Pang et al. (2025). Each coloured ribbon displays the estimate obtained using data in Figure 1 up to the day the coloured ribbon ends (i.e. black ribbon uses only data until the time of the intervention; various colours use more and more data; and light grey uses the full data until end of September, corresponding to the fit shown in Figure 1). The number of knots used is obtained flooring the number of days used divided by 16, to maintain the gap between knots roughly constant at two generation times (here, 3-day latent + 5-day infectious period). Black horizontal lines give the actual R value used in the simulation in Figure 1. Note how using more data in the future affects the R estimates also in the past—the used data are treated as a static set of points, with no underlying concept of time. Also, note the GAM tendency to produce oscillations due to the nature of the spline fit around sudden changes; in particular, estimates of R above 1.5 in early July (e.g. light grey, but other colours too) are *not* driven by a data up to or around that time, but rather by the presence of data *well after* the change (especially late August and September).

brought all regions to the same high prevalence, forcing Lockdown 2 (November). With a short Lockdown 2 and Alpha present, Lockdown 3 was also inevitable (plus, at that point aimed at securing widespread vaccination).

- The goal behind decisions is not always clearly specified, and conflicting priorities emerge over time. Also, minimizing overall deaths is sensible, but not overwhelming hospitals limits *avoidable* deaths (more unpalatable, involving younger people, etc.). Furthermore, the cost of a collapsed health-care system is incalculable.
- Principle of precaution: would the authors let admissions shoot across hospital capacity gambling they will soon turn around because they suspect low herd immunity or spontaneous social distancing? What if shielding of elderlies was modelled/parameterized imperfectly or experiences implementation challenges? Remember that so much was learnt about hospital/ICU case management within just the first few pandemic months.

*Conflicts of interest:* Member of SPI-M-O, the Scientific Pandemic Influenza group on Modelling which became Operational during the pandemic (March 2020–March 2022). Not member of SPI-M since. Seconded to the UK Health Security Agency November 2021-March 2023 (honorary appointment since).

Code available at: 10.5281/zenodo.16267332

## References

Pang, X., Han, Y., Tressier, E., Abdul Aziz, N., Pellis, L., House, T., & Hall, I. (2025). Time-varying reproduction number estimation: Fusing compartmental models with generalized additive models. *Journal of the Royal Society, Interface*, *22*(222), 20240518. https://doi.org/10.1098/rsif.2024.0518

Pellis, L., Scarabel, F., Stage, H. B., Overton, C. E., Chappell, L. H., Fearon, E., Bennett, E., Lythgoe, K. A., House, T. A., Hall, I., & University of Manchester COVID-19 Modelling Group. (2021). Challenges in control of COVID-19: Short doubling time and long delay to effect of interventions. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200264. https://doi.org/10.1098/rstb.2020.0264

https://doi.org/10.1093/jrsssa/qnaf117
Advance access publication 18 July 2025

# Alex Selby's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Alex Selby**

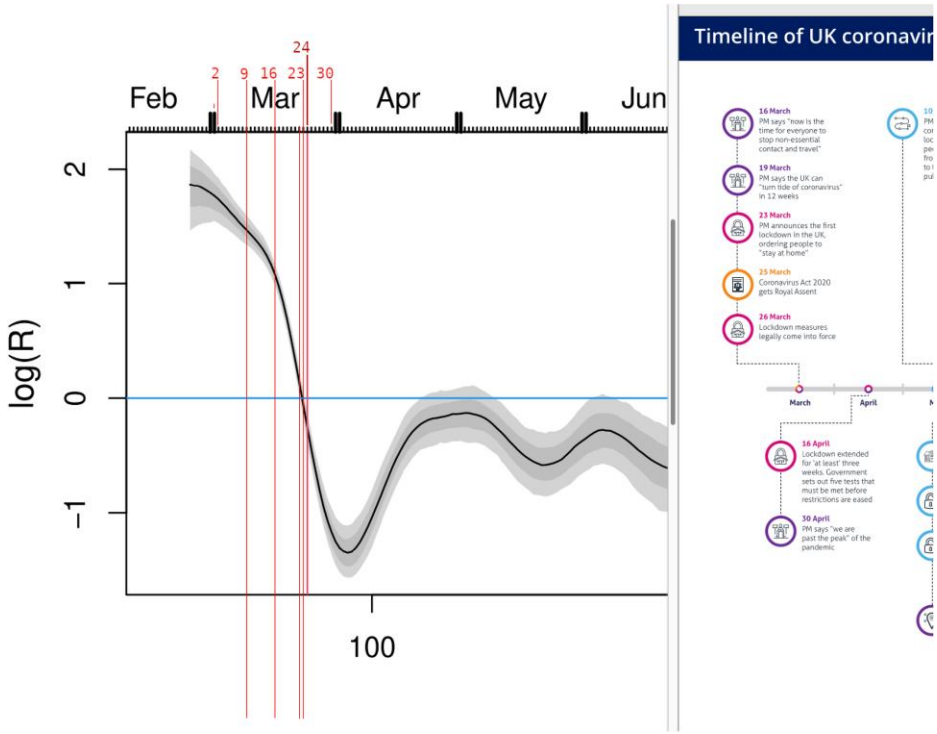
Independent Researcher, Cambridge, UK

*Address for correspondence*: Email: alex.selby@pobox.com

Wood et al. (2005) calculated $R < 1$ before lockdown from which the case is made that a full lockdown was 'largely unnecessary'. What follows is where I believe this argument may be lacking. (This note is agnostic about whether lockdown was the best option.)

## 1. Precision

The pictured error bands in $R$ in Figure 10 (below) do not account for uncertainties in the infection–death distribution, generation time, infectivity duration, or shortcomings of using a single compartment (care homes, ages/vulnerable, regional variations, etc.).



The graph shows log($R$) crossing zero just fractionally before 23 March, the day of the stay at home order, from which the authors conclude that $R < 1$ on the first *full* day of lockdown, 24 March (dark red line). Given this virtual dead heat and modelling uncertainties, it seems too much to conclude that '$R < 1$ clearly before the first lockdown'.

Figure 10 (above), obtained from Wood et al. (2005), with vertical lines added to show relevant days. Juxtaposition: timeline of UK coronavirus measures.

## 2. Momentum from those infections

The focus is on whether infections are falling, but they also have to fall fast enough. Granting $R = \sim 1$ at lockdown, the result of not locking down would, under simple SIR assumptions, lead to ~100k deaths, though possibly a lot more or a lot less depending on the exact value of $R$ at lockdown, cf. ~35k deaths during the actual first wave.

## 3. Limited headroom

Deaths in England peaked at around 1,000/day on 8 April and only subsided to 100/day by ~8 June. A crude calculation suggests an average post-lockdown $R$ of around (100/1,000)^(5/61) = ~0.83 which means even increasing the contact rate by 10% would result in substantially many more first-wave infections and deaths. See also Johnson (2025).

## 4. Interpretation

It is uncontroversial that contact rates reduced before lockdown(s), but there is a question of whether people were voluntarily reducing contacts (a) in anticipation of a legal lockdown they believed was coming anyway (or already in force) or (b) because they felt the situation was getting bad. It seems plausible that (a) contributes to some extent considering, e.g. that the Prime Minister announced a week before official lockdown, that 'now is the time for everyone to stop non-essential contact and travel' (Prime Minister's statement on coronavirus (COVID-19), 2020), and also talked of compulsory shielding for vulnerable groups. This would mean that in the counterfactual no-lockdown scenario, $R$ should be revised upwards *before* lockdown itself, which would significantly increase the first-wave deaths.

## 5. Arguably misleading emphasis on peak incidence

In the study by Wood et al. (2005), Figure 10 (above), Figure 11, and much of the discussion is around the peak in incidence, but it is the peak in $R$ that really matters, as changes in behaviour change $R$ instantly, and the peak in $R$ lags the peak in incidence. Just after peak incidence, $R$ can be greater than 1, and if $R$ is maintained at that level, incidence can increase again.

## 6. Frontloading infections

Section 5.2 argues that countries with greater first-wave attack rates had reduced second waves. But different countries were different in many ways, and the second wave was largely due to Alpha (not mentioned in the paper) for which the UK was unhappily in the front line, and without which the second wave would have likely been mild, even under lighter restrictions. It was basically bad luck that such a variant hit just as an effective vaccine was about to be rolled out, so it would not have made sense in March 2020 to infect a lot of people to protect against that eventuality. {Initially, problematic variants were considered unlikely due to coronaviruses' error correction mechanisms (Manzanares-Meza & Medina-Contreras, 2020; Moshiri, 2020; NERVTAG: NT-SARS-CoV-2 variants, 2020).}

## 7. Overstated herd immunity

The statement 'In consequence the pattern of increase-lockdown-decrease is simply inevitable and conveys no information about lockdowns' role in reversing waves of infection …' could be misleading. Decrease is only inevitable (in the absence of other measures) when the pool of susceptibles is significantly depleted, and that was far from being true in March 2020: by 23 March, only 2%–3% of the population had been infected (based on an IFR of ~1%).

*Conflicts of interest:* none declared.

## References

Johnson, O. (31 March 2025). *Johnson's contribution to the Discussion of 'Statistical aspects of the Covid-19 response' by Wood et al.* arXiv 24308. https://doi.org/10.48550/arXiv.2503.24308, 31 March, 2025, preprint: not peer reviewed.

Manzanares-Meza, L. D., & Medina-Contreras, O. (2020). SARS-CoV-2 and influenza: A comparative overview and treatment implications. *Boletin Medico Del Hospital Infantil De Mexico*, *77*(5), 262–273. https://doi.org/10.24875/BMHIM.20000183

Moshiri, N. (6 April 2020). Coronavirus seems to mutate much slower than seasonal flu. Science news article. https://www.livescience.com/coronavirus-mutation-rate.html

NERVTAG: NT-SARS-CoV-2 variants (13 May 2020). https://www.gov.uk/government/publications/nervtag-nt-sars-cov-2-variants-13-may-2020

Prime Minister's statement on coronavirus (COVID-19) (16 March 2020). https://www.gov.uk/government/speeches/pm-statement-on-coronavirus-16-march-2020

Wood, S. N., Wit, E. C., McKeigue, P. M., Hu, D., Flood, B., Corcoran, L., & Abou Jawad, T. (2025). Some statistical aspects of the Covid-19 response. *Journal of the Royal Statistical Society Series A: Statistics in Society (to appear)*.

https://doi.org/10.1093/jrsssa/qnaf095
Advance access publication 30 June 2025



# Stephen Senn's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Stephen Senn**

Consultant Statistician, Edinburgh, UK

*Address for correspondence*: Stephen Senn, Consultant Statistician, 29 Merchiston Crescent, Edinburgh EH10 5AJ, UK. Email: stephen@senns.uk

> "lo and behold, the preceding edition of the chromate file card bore the direction to add '2 or 3' drops not '23'. Primo Levi, Chromium, *The Periodic Table*, (p131). (Levi, 2000)

My first job was as Medical Information Officer for the Tunbridge Wells Health District, in which role I learned many things, most important of all, the depressing lesson that if a statistic looks interesting, it is almost certainly wrong. Notwithstanding this lesson, I am sure that many interesting things in this splendid and comprehensive analysis by Wood et al. are right. However, one in particular piqued my attention and I suspect that it *is* wrong.

Benjamin Franklin's two certainties, death and taxes are surely linked to income but whereas in the UK, the relationship of the marginal rate of tax to income is not monotonic, a reflection no doubt of the fact that our legislators are, as a class, numerically challenged, I find it hard to believe in a non-monotonic relationship between life expectancy and deprivation. Mother Nature has no difficulty with science, even if it is a closed book to our rulers, who must have found the pandemic particularly challenging.

The left-hand panel of Figure 3 looks wrong to me. It suggests that for males but not for females, there is a sweet spot of just a little deprivation, with the 7th tenth outperforming not only tenths one to six but also tenths eight and nine and matching tenth number ten. I decided to go

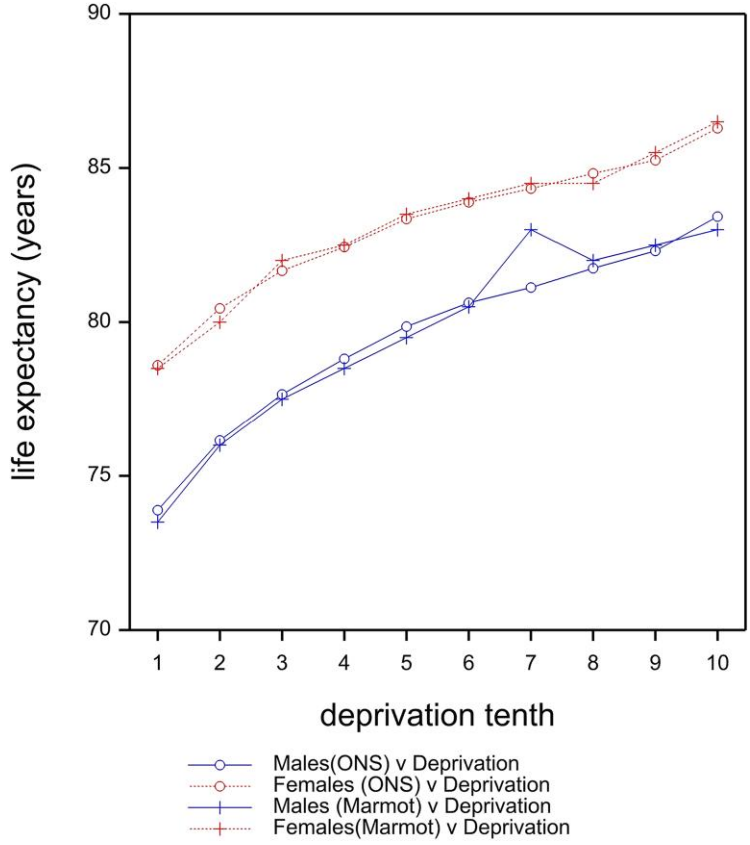



**Figure 1.** Data on life expectancy by deprivation tenth. Open circles, o, are from downloaded ONS figures (Office for National Statistics, 2020) and pluses, +, obtained from the report by Marmot et al. (2020).

back to the Marmot report (Marmot et al., 2020) and then in turn to the ONS figures (Office for National Statistics, 2020). What I found is plotted in the Figure 1 below. The Marmot values were read off from a graph using a ruler and the ONS ones calculated from a downloaded spreadsheet. I could easily have a made a mistake in either. However, the Marmot Male figure for the 7th tenth looks like a bit of a 'Trout in the Milk' (Wainer, 2005) to me. I suspect a transcription error.

Of course, this sort of thing makes one worry about other figures. The argument regarding the right hand panel of Figure 3 strikes me in any case as being rather tenuous, based as it is on ratios of derivatives. One would like to be reassured, as a minimum, that the figures are correct. Can the authors provide such an assurance?

The serious point of all this is that the standard advice (which I am sure I have frequently been guilty of ignoring) is sound: when quoting figures, go back to the source.

Finally, if I may make a pedantic point (what else should one expect from a retired statistician?). The nine cuts that divide a distribution into tenths are usefully described as *deciles*. To label the ten *tenths* themselves as *deciles*, makes the term ambiguous, substitutes a less common two syllable one for a simple one syllable word we all understand and makes no more sense than calling the two halves of a distribution the medians. Let's not.

*Conflicts of interest:* none declared.

## References

Levi, P. (2000). *The periodic table*. Penguin.

Marmot, M., Allen, J., Boyce, T., Goldblatt, P., & Morrison, J. (2020). *Marmot review 10 years on*. Institute of Health Equity. https://www.instituteofhealthequity.org/resources-reports/marmot-review-10-years

Office for National Statistics. (2020). *Health state life expectancies by national deprivation deciles, England: 2016 to 2018 Life expectancy and years expected to live in "Good" health using national indices of deprivation to measure socioeconomic inequalities in England*. Office for National Statistics.
Wainer, H. (2005). *Graphic discovery: A trout in the milk and other visual adventures*. Princeton University Press.

https://doi.org/10.1093/jrsssa/qnaf068
Advance access publication 26 June 2025



# Garib Nath Singh's contribution to the Discussion of 'Some statistical aspects of the COVID-19 Response' by Wood et al.

**Garib Nath Singh**

Department of Mathematics and Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad 826004, India

*Address for correspondence*: Garib Nath Singh, Department of Mathematics and Computing, Indian Institute of Technology (Indian School of Mines), Dhanbad 826004, Jharkhand, India. Email: gnsingh@iitism.ac.in

In this manuscript, the authors have discussed issues pertaining to the statistical representation of the UK's response to the COVID-19 pandemic, with the goal of understanding how biased or misleading statistical approaches affects public perception and government policy.

The manuscript rightly criticizes the statistical methodologies used in studies on long COVID, particularly issues like selection bias and lack of control groups. However, it tends to overlook the challenges involved in studying a novel disease under pandemic conditions. Initial research frequently relies on imperfect data, since large-scale, controlled studies take time. Additionally, the manuscript suggests that the definition of long COVID suffers from the assumption that symptoms are caused by COVID. However, persistent symptoms that follow a specific infection have epidemiological significance. The manuscript is also critical of self-reported data, but has not suggested suitable alternatives.

While inferring the existence of a causal link between economic shocks and life expectancy reduction, the study oversimplifies the complex nature of these relationships and fails to take a nuanced approach involving other confounding factors. The study also oversimplifies the impact of the pandemic response on excess deaths and may misattribute some trends to demographic aging instead of direct pandemic effects.

The manuscript points out the dangers of polarization in online discourse as well as the hindrance caused by the adversarial nature of some scientific debates. Discussions on ethical and practical implications of lockdowns is of utmost importance. There is a scope for elaboration on these matters, and suggesting remedial measures for the same. There is also a need for empirical comparisons to illustrate how the theoretical refinements improve predictive accuracy.

Striking a balance between caution and urgency in crisis is a challenge. Some shortcomings become more obvious in hindsight. The manuscript highlights important lessons for future public health emergencies, with an emphasis on transparency, improved communication and a balanced approach to scientific debate.

*Conflicts of interest:* None declared.

https://doi.org/10.1093/jrsssa/qnaf082
Advance access publication 27 June 2025

# David Spiegelhalter's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**David Spiegelhalter**

Department of Statistics, University of Cambridge, Cambridge, UK

*Address for correspondence*: David Spiegelhalter, Department of Statistics, University of Cambridge, Cambridge CB3 0WB. Email: david@statslab.cam.ac.uk



I welcome the opportunity for the Society to discuss statistical issues arising from the COVID-19 pandemic, but I can only cover a few topics in the space allowed. However, for further discussion, see Spiegelhalter and Masters (2021).

The authors rightly emphasize the huge age-gradient in risk from infection with COVID-19. It was clear from March 2020 that this risk was remarkably proportional to background actuarial risks, as if COVID magnified any existing risk—the similarity of risk factors for dying from COVID-19 to 'standard' risk factors was later confirmed (Bhaskaran et al., 2021).

But the government appeared very reluctant to make this clear, and instead issued generic fear-mongering advertisements. From my personal experience, government agencies actively discouraged the development of a public-facing tool for clearly communicating risk at an individual level, in spite of research on the appropriate ways of doing this (Freeman et al., 2021). One of my many regrets of the pandemic is that we did not independently produce and promote a personalized risk tool.

With regard to excess deaths, Office for National Statistics (ONS) did use a fairly basic method in their weekly bulletins on death registration, which was changed in early 2024 to a more sophisticated model (Office for National Statistics, 2024). However, other ONS publications in international comparisons appropriately used age-standardized mortality rates (ASMRs) (Office for National Statistics, 2022a).

The authors' Figure 15 shows observed hospital occupancy being far lower than modelled projections. The government decided against lockdown at Christmas 2021, even though Omicron infections had been over 400,000 a day near the end of December (Office for National Statistics, 2022b). But behaviours over the high-risk Christmas period may have been influenced by Chris Whitty's widely-reported advice to the public to 'prioritize social interactions that really matter to them' and 'don't mix with people you don't have to' (Stewart et al., 2021). This reveals the problem with models where the most important component—human behaviour—is the one that is known least about. Calculated uncertainty intervals are inadequate to express this source of uncertainty, and in her 2022 RSS Presidential address, Sylvia Richardson pointed out that 'The continual back-loop between behavioural changes and epidemic evolution made all but very short-term predictions somewhat futile' (Richardson, 2022).

An important lesson for future pandemics might be for modellers to explore a full range of possibilities rather than focus on traditional government concern for reasonable worst-case scenarios—of course these may lead to a huge range in possible futures, and subsequent accusations of being unhelpful, but that may be a better reflection of realistic uncertainties.

*Conflicts of interest:* None declared.

## References

Bhaskaran, K., Bacon, S., Evans, S. J., Bates, C. J., Rentsch, C. T., MacKenna, B., Tomlinson, L., Walker, A. J., Schultze, A., Morton, C. E., Grint, D., Mehrkar, A., Eggo, R. M., Inglesby, P., Douglas, I. J., McDonald, H. I., Cockburn, J., Williamson, E. J., Evans, D., … Goldacre, B. (2021). Factors associated with deaths due to COVID-19 versus other causes: Population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. *The Lancet Regional Health: Europe*, *6*, 100109. https://doi.org/10.1016/j.lanepe.2021.100109

Freeman, A. L., Kerr, J., Recchia, G., Schneider, C. R., Lawrence, A. C., Finikarides, L., Luoni, G., Dryhurst, S., & Spiegelhalter, D. (2021). Communicating personalized risks from COVID-19: Guidelines from an empirical study. *Royal Society Open Science*, *8*(4), 201721. https://doi.org/10.1098/rsos.201721

Office for National Statistics. (2022a). *Comparisons of all-cause mortality between European countries and regions*. https://www.ons.gov.uk/releases/internationalcomparisonsofallcauseexcessmortalityineuropeannationsandregions. Date accessed July 2025.

Office for National Statistics. (2022b). *Coronavirus (COVID-19) Infection Survey: England*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/coronaviruscovid19infectionsurveydata/2022. Date accessed July 2025.

Office for National Statistics. (2024). *Estimating excess deaths in the UK, methodology changes*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/estimatingexcessdeathsintheukmethodologychanges/february2024. Date accessed July 2025.

Richardson, S. (2022). Statistics in times of increasing uncertainty. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(4), 1471–1496. https://doi.org/10.1111/rssa.12957

Spiegelhalter, D., & Masters, A. (2021). *COVID by numbers: Making sense of the pandemic with data*. Penguin UK.

Stewart, H., Allegretti, A., Davis, N., & Gregory, A. (2021, December 15). Cut back on socialising, says Whitty, as he predicts surge of hospitalisations. *The Guardian*. https://www.theguardian.com/world/2021/dec/15/uk-covid-cases-at-record-level-with-78610-new-infections. Date accessed 2025.

https://doi.org/10.1093/jrsssa/qnaf105
Advance access publication 1 July 2025

# Andrej Srakar's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Andrej Srakar**

Institute for Economic Research and University of Ljubljana, Ljubljana, Slovenia

*Address for correspondence*: Andrej Srakar, Institute for Economic Research and University of Ljubljana, Kardeljeva pl. 17, 1000 Ljubljana, Slovenia. Email: srakara@ier.si

I congratulate the authors for their excellent and mind-provoking paper! My short remarks point to theoretical and statistical modelling considerations regarding the Covid-19 pandemic, which the authors shortly address, as well as more general data availability aspects.

Firstly, modelling of Covid-19 pandemic faced clear problems of estimating causality. Lockdowns were of endogenous nature which often went neglected. While one of the paper authors, Simon N. Wood pointed to this in his 2022 RSS Conference plenary lecture, this issue is not addressed in the paper. It would be great to learn how biased have been causal assessments of the effects of lockdowns due to problems in causality. Another causal issue is the one that the authors mention in the paper, nonrandomized nature of the prevalence and similar Covid-19 assessments. Issues of Simpson's paradox related to unobserved confoundedness have also emerged during the pandemic time. It would be great to learn if causal inference developments have contributed to better statistics about Covid-19.

Authors also address SEIR-type compartmental models. They relate to probability theory where modelling of infectious diseases is an active research area. I wonder whether the pandemic has contributed to some significant novelties in either theoretical or applied probability in the modelling of infectious diseases (say, advances in contact and branching processes). Can and did Covid-19 pandemic experiences contributed to modelling randomness using mathematics in any way, either

standalone or in combination with more general statistical approaches from, for example, Bayesian, causal, high-dimensional, network, or similar approaches?

Data availability proved a big issue in times of the Covid-19 pandemic. As the authors write, the public availability of data was in many respects exemplary (e.g. ONS, ISARIC, REACT) but other data were effectively closed to general scrutiny, such as data on nosocomial infection. I wonder how to improve data capacities and which ones to improve in preparedness for any similar future catastrophes and societal shocks (in health, climate change, economy, and other areas). Do we need more and better, open data? And how to ensure preparedness in terms of statistics in general terms for similar future societal shocks? Are open data a good response here? Or do we need higher investments in methodological developments and modelling centres? Or both, or anything else?

I hope my comments would be a good contribution to the discussion of an important paper.

*Conflicts of interest:* none declared.

https://doi.org/10.1093/jrsssa/qnaf119
Advance access publication 16 July 2025



# StatGroup-19's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Pierfrancesco Alaimo Di Loro[1], Fabio Divino[2], Alessio Farcomeni[3], Giovanna Jona Lasinio[4], Gianfranco Lovison[5], Antonello Maruotti[1] and Marco Mingione[6]**

[1]Department of GEPLI, LUMSA University, Via Pompeo Magno 28, Rome, 00192, Italy
[2]Department of Biosciences and Territory, University of Molise, Isernia, Italy
[3]Department of Economics and Finance, University of Rome "Tor Vergata", Rome, Italy
[4]Department of Statistical Sciences, Sapienza University of Rome, Piazzale A. Moro 5, 00185, Rome, Italy
[5]Department of Public Health, University of Basel, Basel, Switzerland
[6]Department of Political Sciences, Roma Tre University, Via G. Chiabrera 199, 00145 Roma, Italy

*Address for correspondence*: Marco Mingione, Department of Political Sciences, Roma Tre University, Via G. Chiabrera 199, 00145 Roma, Italy. Email: marco.mingione@uniroma3.it

First, we would like to commend the authors for their insightful and comprehensive paper, which further highlights the widespread use of statistically questionable approaches in response to the COVID-19 pandemic. The article raises numerous important discussion points, and we wish to contribute by highlighting one crucial aspect that affected the unfolding of the Italian pandemic: the distorted communication of epidemic figures by media and politics, resulting in the misperception of risk among the general public. This issue is pivotal in the context of crisis management, and similar patterns emerged in many other countries.

We worked together ('StatGroup-19', for the media) since the start of the pandemic, as the crisis required rigorous statistical expertise. Our objectives were twofold: develop robust tools to forecast the progression of the epidemic (Alaimo Di Loro et al., 2024, 2021; Bartolucci & Farcomeni, 2022; Farcomeni et al., 2021; Mingione et al., 2021, 2024; Sherratt et al., 2023), and counter the misleading narratives disseminated to the general public. The pandemic made it clear to us that lack of statistical literacy severely undermined media communication and political decision-making. These issues have been discussed in position papers (Alaimo Di

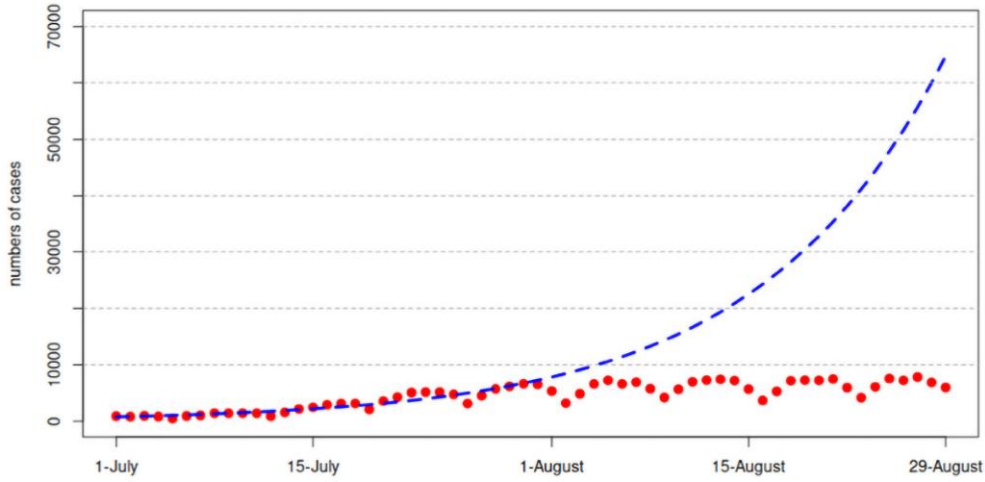



**Figure 1.** COVID-19 in Italy: Daily incidence cases in July and August 2021 (points) versus the Poisson exponential curve estimated using the training cases observed from 1.7 to 31.7 (dashed line). Source: Divino, Ciccozzi, et al. (2022).

Loro et al., 2022; Divino, Ciccozzi, et al., 2022; Divino, Maruotti, et al., 2022; Jona Lasinio et al., 2022).

Risk perception plays a fundamental role in shaping the course of any phenomenon influenced by human behaviour, be it epidemiological or otherwise. From the outset of the pandemic, the media thrived on spreading scaremongering messages, routinely speculating about new restrictions ahead of public holidays or in response to even the slightest increase in case numbers. Meanwhile, politics transformed what should have been a scientific issue into a battleground for advancing personal and party agendas. These narratives were sometimes supported by distinguished scholars and scientists from various fields, some of whom even endorsed unscientific claims (such as 'long-term exponential growths', Figure 1).

In Italy, significant reliance for decision-making was placed on delayed estimates of catchy and easily communicated metrics, such as the reproduction number $R_t$. However, the limitations of such metrics were rarely discussed (Gostic et al., 2020; Ioannidis et al., 2020), and in practice ignored.

The public often lived in a state of persistent confusion. In the absence of clear and consistent messages, rational discourse gave way to partisan alignment. As a result, public trust in institutions and medical advice from public health authorities declined, ultimately reducing cooperation at precisely the time when it was most needed.

*Conflicts of interest:* None declared.

## References

Alaimo Di Loro P., Böhning D., & Sahu S. K. (2024). A Bayesian spatio-temporal Poisson auto-regressive model for the disease infection rate: Application to covid-19 cases in England. *Journal of the Royal Statistical Society: Series C, Applied Statistics*, *74*(3), 551–575. https://doi.org/10.1093/jrsssc/qlae067

Alaimo Di Loro P., Divino F., Farcomeni A., Jona Lasinio G., Lovison G., Maruotti A., & Mingione M. (2021). Nowcasting Covid-19 incidence indicators during the Italian first outbreak. *Statistics in Medicine*, *40*(16), 3843–3864. https://doi.org/10.1002/sim.v40.16

Alaimo Di Loro P., Divino F., Farcomeni A., Jona Lasinio G., Lovison G., Maruotti A., & Mingione M. (2022). Covid-19 in Italy: Modelling, communications, and collaborations. *Significance*, *19*(2), 19–21. https://doi.org/10.1111/1740-9713.01629

Bartolucci F., & Farcomeni A. (2022). A spatio-temporal model based on discrete latent variables for the analysis of COVID-19 incidence. *Spatial Statistics*, *49*, 100504. https://doi.org/10.1016/j.spasta.2021.100504

Divino F., Ciccozzi M., Farcomeni A., Jona-Lasinio G., Lovison G., & Maruotti A. (2022). Unreliable predictions about COVID-19 infections and hospitalizations make people worry: The case of Italy. *Journal of Medical Virology*, *94*(1), 26–28. https://doi.org/10.1002/jmv.v94.1

Divino F., Maruotti A., Farcomeni A., Jona-Lasinio G., Lovison G., & Ciccozzi M. (2022). On the severity of COVID-19 infections in 2021 in Italy. *Journal of Medical Virology*, *94*(4), 1281–1283. https://doi.org/10.1002/jmv.v94.4

Farcomeni A., Maruotti A., Divino F., Jona-Lasinio G., & Lovison G. (2021). An ensemble approach to short-term forecast of COVID-19 intensive care occupancy in Italian regions. *Biometrical Journal*, *63*(3), 503–513. https://doi.org/10.1002/bimj.v63.3

Gostic K. M., McGough L., Baskerville E. B., Abbott S., Joshi K., Tedijanto C., Kahn R., Niehus R., Hay J. A., De Salazar P. M., Hellewell J., Meakin S., Munday J. D., Bosse N. I., Sherrat K., Thompson R. N., White L. F., Huisman J. S., Scire J., …Cobey S. (2020). Practical considerations for measuring the effective reproductive number, Rt. *PLoS Computational Biology*, *16*(12), 1–21. https://doi.org/10.1371/journal.pcbi.1008409

Ioannidis J. P. A., Cripps S., & Tanner M. A. (2020). Forecasting for COVID-19 has failed. *International Journal of Forecasting*, *38*(2), 423–438. https://doi.org/10.1016/j.ijforecast.2020.08.004

Jona Lasinio G., Divino F., Lovison G., Mingione M., Alaimo Di Loro P., Farcomeni A., & Maruotti A. (2022). Two years of COVID-19 pandemic: The Italian experience of statgroup-19. *Environmetrics*, *33*(8), e2768. https://doi.org/10.1002/env.v33.8

Mingione M., Alaimo Di Loro P., Farcomeni A., Divino F., Lovison G., Maruotti A., & Jona Lasinio G. (2021). Spatio-temporal modelling of COVID-19 incident cases using richards' curve: An application to the Italian regions. *Spatial Statistics*, *49*(16), 100544. https://doi.org/10.1016/j.spasta.2021.100544

Mingione M., Alaimo Di Loro P., & Maruotti A. (2024). A useful parametric specification to model epidemiological data: Revival of the richards' curve. *Statistical Methods in Medical Research*, *33*(8), 1473–1494. https://doi.org/10.1177/09622802241262522

Sherratt K., Gruson H., Grah R., Johnson H., Niehus R., Prasse B., Sandmann F., Deuschel J., Wolffram D., Abbott S., Ullrich A., Gibson G., Ray E. L., Reich N. G., Sheldon D., Wang Y., Wattanachit N., Wang L., Trnka J., … Funk S. (2023). Predictive performance of multi-model ensemble forecasts of COVID-19 across European nations. *eLife*, *12*, e81916. https://doi.org/10.7554/eLife.81916

https://doi.org/10.1093/jrsssa/qnaf081
Advance access publication 27 June 2025

# Yu Tian, Tan Meng, Maozai Tian, and Zhaoya Zhou's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Yu Tian[1], Tan Meng[2], Zhaoya Zhou[2] and Maozai Tian[2,3,4]**

[1]School of Marxism, Xinjiang University of China, Urumqi 830046, China
[2]Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China
[3]School of Statistics and Data Science, Xinjiang University of Finance and Economics, Urumqi 830012, China
[4]Department of Mathematics and Data Science, Changji University, Changji 831100, China

*Address for correspondence*: Maozai Tian, Center for Applied Statistics, School of Statistics, Renmin University of China, 100872, Beijing, China. Email: mztian@ruc.edu.cn

## 1 Comments and suggestions

We congratulate Wood et al. for discussing some statistical aspects of the UK Covid-19 pandemic response. This article gives an in-depth discussion on the statistical applications, misuse and abuse in the fight against the epidemic. To one who has worked extensively on the pandemic response

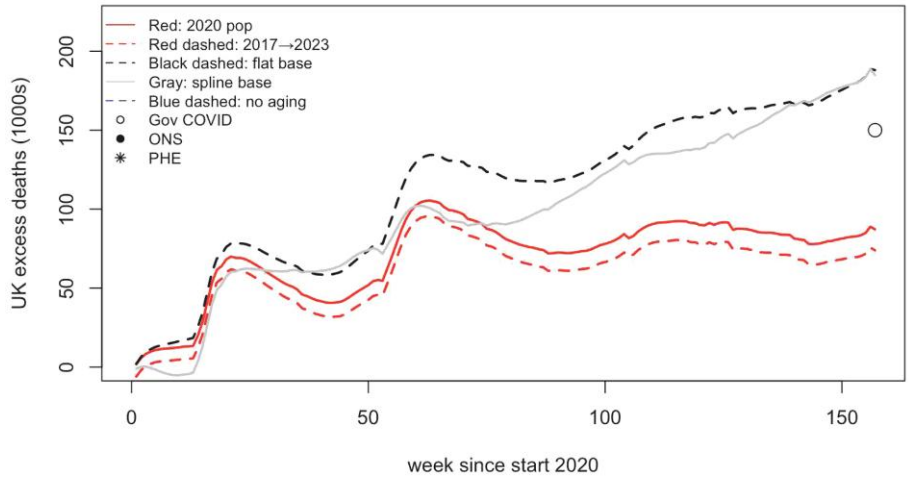



**Figure 1.** Red solid is the cumulative excess deaths from the start of 2020 iterating from population by age data for the start of 2020; red dashed is the equivalent iterating 6 years from population by age data for the start of 2017; black dashed is for the conventional method based on 2017–19 data; grey is the same but using life table iterated weekly deaths in place of raw deaths; blue dash is the life table based approach with ageing turned off, starting from the estimated population by age in mid-2018.

and related areas, this is all a new mirror image universe. It can be sure that the same will apply to many readers of journal JRSS-A and may inspire many follow-up works.

Covid-19 pandemic has historically posed profound threats to human being worldwide, with claiming millions of lives and exposing the fragility of global health systems. The emergence of the COVID-19 in late 2019 escalated this threat to an unprecedented scale. This article explores several statistical aspects of the response to the COVID-19 in the UK, paying special attention to the situations where statistically questionable methods or presentation has had a substantial impact on public perception, or government policy. Some issues regarding the use of life loss data, excess mortality, and case data are raised. The consequences of ignoring most individual differences in epidemic models, and some other statistically significant omissions are covered. All the discussions of the article have important practical value. Here are several comments and suggestions:

1. **Wide calibre and narrow calibre**. Unfortunately, this article seriously overlooked the issue of inconsistent sampling calibres. The definition of 'wide calibre' here means that as long as COVID-19 can be detected after death, it will be included in COVID-19 deaths. In contrast, some countries adopt a 'narrow calibre' approach. This definition requires that death must be caused by pneumonia or respiratory failure directly caused by COVID-19. If the death is caused by other diseases or basic diseases, it will not be regarded as COVID-19's death. This difference in statistical calibre may lead to differences in the death data of COVID-19 among countries. Understanding these differences will help us to assess the impact of the epidemic more comprehensively. See, e.g. Tian (2025a, 2025b), Huang et al. (2020), Wang et al. (2020), Liang et al. (2022), Rui & Tian (2021), Rui et al. (2021), Ma & Tian (2025), Rui et al. (2025), Yu et al. (2022), Tian & Yu (2024), and references therein.
2. **Statistical model inference**. For the sake of rigorous writing and ease of reading, it is recommended to list the main models or mathematical formulas used in the article or attach them. Empirically, Figure 1 depicts the red solid being the cumulative excess deaths from the start of 2020 iterating from population by age data for the start of 2020; red dashed the equivalent iterating 6 years from population by age data for the start of 2017; Figure 1 of around 95,000 total excess deaths from the solid red curve is lower than the figure of 167,356 given by the ONS as the excess death figure from March 2020 until the end of 2022, or the PHE figure for England of 144,446. Let $y_i$ denote the deaths in week $w_i$ of the year, corresponding to time $t_i$ since the start of the data. To estimate the annual cycle in death rates, the generalized

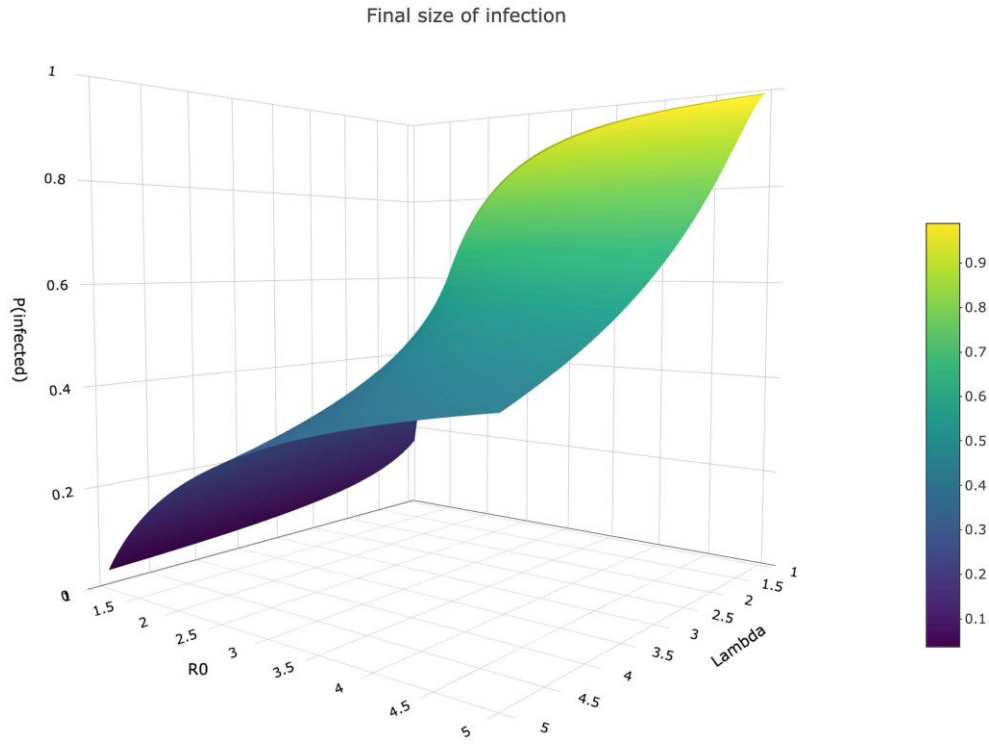



**Figure 2.** The proportion infected against $\lambda$ for $R_0 = 2, 3, 4, 5$ (ascending, red, green, blue, turquoise), for SEIR epidemic models with susceptibility or mixing rates varying across individuals.

additive model

$$\mu_i = f_1(w_i) + f_2(t_i), \qquad \frac{y_i - \mu_i}{\sigma} \sim t_v.$$

Note that the above model $\frac{y_i-\mu_i}{\sigma} \sim t_v$ is incorrect. It should be approximately $N(0, 1)$, the standard normal distribution. Another question here is how to estimate the two functions $f_1(w_i)$ and $f_2(t_i)$?, where $f_1$ is a cyclic smooth function and $f_2$ is a centred slowly varying smooth function, and how to determine the number of iterations of ageing and deaths?

3. **Optimal parametric estimation**. For many substantial infections, there are significant incubation periods during which individuals have been infected but are not yet infectious themselves. Consider the standard SEIR model for this situation is,

$$\begin{aligned}
\frac{ds(\alpha, t)}{dt} &= -\alpha s(\alpha, t) I(t) \\
\frac{de(\alpha, t)}{dt} &= \alpha s(\alpha, t) I(t) - \delta e(\alpha, t) \\
\frac{di(\alpha, t)}{dt} &= \delta e(\alpha, t) - \gamma i(\alpha, t),
\end{aligned}$$

where $I(t) = \int i(\alpha, t) d\alpha$. For the above person-to-person variability in SEIR models, how to get the optimal estimators of the parameters?

Figure 2 presents the age band dependent variability used by SAGE models in the UK corresponding to empirical $\lambda \approx 1.2$, refer to Figure 2. Realistic estimates seem to be in the range 2.5–5. The question here is there any criteria to give automatically optimal estimators of the parameters $\lambda$ and the basic reproduction number $R_0$? The study has generated several weights by thumb. Theoretically, the validity of the weighting procedure is needed.

## Acknowledgments

The authors thank the RSS Journals Manager Judith Shorten for her kind invitation to make presentation with discussion of the paper. Prof. Maozai Tian's work was supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (25XNN015).

*Conflicts of interest:* None declared.

## References

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., … Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*, *395*(10223), 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5

Liang, J. W., Zhang, X. L., Wang, K., Tang, M. L. and Tian, M. Z. (2022). Discovering Dynamic Models of COVID-19 Transmission. *Transboundary and Emerging Diseases*, *69*, 64–70. https://doi.org/10.1111/tbed.14263

Ma, S. P., & Tian, M. Z. (2025). A censored quantile transformation model for Alzheimer's disease data with multiple functional covariates. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *188*(2), 515–538. https://doi.org/10.1093/jrsssa/qnae061

Rui, R. X., & Tian, M. Z. (2021). Joint estimation of case fatality rate of COVID-19 and power of quarantine strategy performed in Wuhan, China. *Biometrical Journal*, *63*(1), 46–58. https://doi.org/10.1002/bimj.v63.1

Rui, R. X., Tian, M. Z., Tang, M. L., Ho, T. S., & Wu, C. H. (2021). Analysis of the spread of COVID-19 in the USA with spatio-temporal multivariate time series model. *International Journal of Environmental Research and Public Health*, *18*(2), 774. https://doi.org/10.3390/ijerph18020774

Rui, R. X., Xiong, W., Pan, J. X. and Tian, M. Z. (2025). Unveiling Schizophrenia: A Study with Generalized Functional Linear Mixed Model via the Investigation of Random Functional Effects, *Biostatistics*, *26*(1), 1–22. https://doi.org/10.1093/biostatistics/kxae049

Tian, M. Z., & Yu, K. M. (2024). Maozai Tian and Keming Yu's contributions to the discussion of 'Safe Testing' by Peter Grünwald, Rianne de Heide, and Wouter Koolen. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *86*(5), 1160. https://doi.org/10.1093/jrsssb/qkae076

Tian, M. Z. (2025a). Maozai Tian's Contribution to the Discussion of the 'Discussion Meeting on the Analysis of Citizen Science Data', *Journal of the Royal Statistical Society Series A: Statistics in Society*, *188*(3), 720–722. https://doi.org/10.1093/jrsssa/qnaf021

Tian, M. Z. (2025b). Maozai Tian's Contribution to the Discussion of 'Methods for Estimating the Exposure-Response Curve to Inform the New Safety Standards for Fine Particulate Matter' by Cork, et al, *Journal of the Royal Statistical Society Series A: Statistics in Society*, qnaf048, 1–3. https://doi.org/10.1093/jrsssa/qnaf048

Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X., & Peng, Z. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA*, *323*(11), 1061–1069. https://doi.org/10.1001/jama.2020.1585

Yu, Z., Yu, K. M., Härdle, W. K, Zhang, X. L., Wang, K., & Tian, M. Z. (2022). Bayesian spatio-temporal modeling for inpatient hospital costs of alcohol-related disorder. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *185*(Supplement_2), S644–S667. https://doi.org/10.1111/rssa.12963

https://doi.org/10.1093/jrsssa/qnaf107
Advance access publication 2 July 2025

# Mike Tildesley's contribution to the Discussion of 'Some statistical aspects of the COVID-19 response' by Wood et al.

**Mike Tildesley**

Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Coventry, UK

*Address for correspondence*: Mike Tildesley, Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. Email: M.J.Tildesley@warwick.ac.uk

I provide here perspective comments on Section 5—'Lockdowns'. It is worth stressing that the incidence plotted in Figure 10 (black line) is not the population-level incidence of infection, but is the incidence in those who will later die from the infection. This, therefore, represents a biased sample of infection, skewed towards the elderly and those with underlying health conditions. Given that advice was given to the vulnerable 4 days before the national-lockdown came into effect (COVID-19: guidance on social distancing and for vulnerable people, 2020), and that many vulnerable individuals were well aware of their risks and were already limiting their interactions, it is not surprising that infection in this subsection of the population had already peaked.

However, the vulnerable will still interact with the rest of the population to some degree, and hence limiting infection in the general population is key. Moreover, the assessment of lockdowns in this paper misses the essential point: the lockdown was not a mechanism for simply influencing the peak of the outbreak, but of reducing the area under the curve—reducing the reproductive ratio ($R$) to as low as possible for as long as possible. In Wood's analysis (Figure 10b), all three lockdowns are closely associated with a switch from positive to negative growth rates (as captured by $\log(R)$; or alternatively a switch from $R > 1$ to $R < 1$ as seen in all models in Figure 14).

Given the intense media coverage and public concern about increasing numbers of hospital admissions or deaths, it is unsurprising that some change in behaviour (and hence a change in $R$) occurred before the date of the lockdown. However, it is impossible to forecast the counterfactual—what would have occurred if a lockdown had not been imposed? It is highly likely that this messaging—that a lockdown was not needed in the UK—would have precipitated a relaxation in the behaviour of the population and an increase in mixing, precisely at the time when infection levels were high and needed to be brought under control.

*Conflicts of interest:* none declared.

## Reference

COVID-19: guidance on social distancing and for vulnerable people. (2020). https://www.gov.uk/government/publications/covid-19-guidance-on-social-distancing-and-for-vulnerable-people

https://doi.org/10.1093/jrsssa/qnaf074
Advance access publication 26 June 2025

# Ondřej Vencálek's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Ondřej Vencálek**

Department of Mathematical Analysis and Application of Mathematics, Faculty of Science, Palacký University in Olomouc, Olomouc, Czech Republic

*Address for correspondence*: Ondřej Vencálek, Department of Mathematical Analysis and Application of Mathematics, Faculty of Science, Palacký University in Olomouc, 17. listopadu 12, 771 46 Olomouc, Czech Republic. Email: ondrej.vencalek@upol.cz

**Keywords:** Covid-19, ONS data, vaccine effectiveness

The current paper by Wood et al. is a must-read for anyone who has ever spoken publicly about COVID-19. It is written with the utmost effort to be objective. The procedure suggested in Section 2.4 could provide insight into an issue not discussed in the article, namely the evaluation of vaccine effectiveness.

Let us look at ONS 'Deaths by vaccination status in England' (Deaths by vaccination status, England). The data file includes this warning: '…the characteristics of people in the different vaccination status groups, such as health, may differ…'

An important pattern can be observed in the non-COVID death rates: Once each wave of the vaccination campaign begins, the group of already vaccinated individuals splits into those who opt for the new dose and those who stay with the previous dose. While non-COVID mortality rate decreases in the former group, it increases in the latter. For example, in the 90+ age group, age-standardized non-COVID death rate in the (very small) group 'first dose, at least 21 days ago' is about 5 times higher than in the group 'second dose, at least 21 days ago', and it is about 3.5 times higher than in the group of unvaccinated in June 2021 (see Figure 1). Does it mean that vaccination not followed by the next dose increases the risk of dying? No, the compared groups differ substantially in their susceptibility to die of non-COVID causes. This fact makes any direct comparison of their COVID-related mortality rates not accounting for this baseline difference highly misleading.

This pattern could result from the decision (or inability) of a part of the frail cohort not to be vaccinated (see Arbel et al., 2023). Different non-COVID death rates indicate differences in the health status of the compared groups, known also as *healthy vaccinee effect* (*HVE*) (see Fürst et al., 2024; Humpherson, 2023; Straka et al., 2024). HVE does not mean that the part of

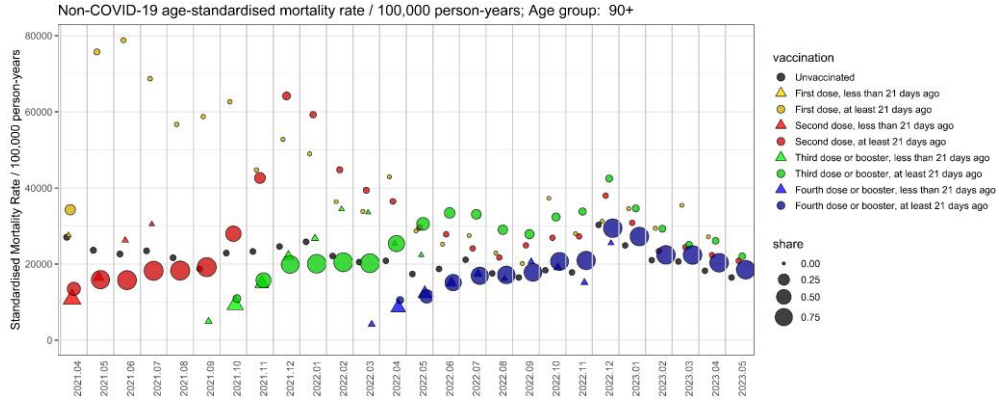



**Figure 1.** Monthly age-standardized mortality rates by vaccination status in the 90+ age group for deaths not involving COVID-19, per 100,000 person-years, England, deaths occurring between 1 April 2021 and 31 May 2023, data (Deaths by vaccination status, England).

population opting for vaccination is 'healthier on average', but it has certain characteristics that make them less likely to die. The presence of HVE can be demonstrated by showing significantly lower non-COVID mortality in a given subgroup than expected in the general population.

The opposite (indication bias, when the frailest are preferentially vaccinated at the beginning of vaccination with the new dose) also plays a role, as can be seen in the 18–39 age group in September 2022.

The effect of the interplay of these two biases on the interpretation of the comparison of death rates in groups defined by vaccination status is surprisingly large.

*Conflicts of interest:* None declared.

## References

Arbel, R., Sergienko, R., Netzer, D., Hammerman, A., Friger, M., Peretz, A., Razi, T., Duskin-Bitan, H., & Yaron, S. (2023). Potential "healthy vaccinee bias" in a study of BNT162b2 vaccine against Covid-19. Reply. *The New England Journal of Medicine*, *389*(3), 284–286. https://doi.org/10.1056/NEJMc2306683

Deaths by vaccination status, England. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/deathsbyvaccinationstatusengland. Date accessed 10 April 2025.

Fürst, T., Bazalová, A., Fryčák, T., & Janošek, J. (2024). Does the healthy vaccinee bias rule them all? Association of COVID-19 vaccination status and all-cause mortality from an analysis of data from 2.2 million individual health records. *International Journal of Infectious Diseases: IJID: Official Publication of the International Society for Infectious Diseases*, *142*, 106976. https://doi.org/10.1016/j.ijid.2024.02.019

Humpherson E To Norman Fenton, Martin Neil, Clare Craig and Scott McLachlan: ONS Deaths by Vaccination Status statistics. (2023). https://osr.statisticsauthority.gov.uk/correspondence/ed-humpherson-to-norman-fenton-martin-neil-clare-craig-and-scott-mclachlan-ons-deaths-by-vaccination-status-statistics/. Date accessed 10 April 2025.

Straka, R., Fürst, T., & Janosek, J. (2024). Healthy vaccinee effect in practice. *Polish Archives of Internal Medicine*, *134*(5), 16750. https://doi.org/10.20452/pamw.16750

https://doi.org/10.1093/jrsssa/qnaf098
Advance access publication 1 July 2025

# Robert Verity, Samir Bhatt, Anne Cori, Seth Flaxman, and Swapnil Mishra's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Robert Verity[1], Anne Cori[1], Swapnil Mishra[2], Seth Flaxman[3] and Samir Bhatt[1,4]**

[1]MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, UK
[2]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
[3]Department of Computer Science, University of Oxford, Oxford, UK
[4]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

*Address for correspondence*: Robert Verity, MRC Centre for Global Infectious Disease Analysis, Imperial College London, London W12 0BZ, UK. Email: r.verity@imperial.ac.uk

During the COVID-19 pandemic, data and analysis took centre stage in shaping public discourse and informing policy decisions. This heightened reliance on modelling placed considerable

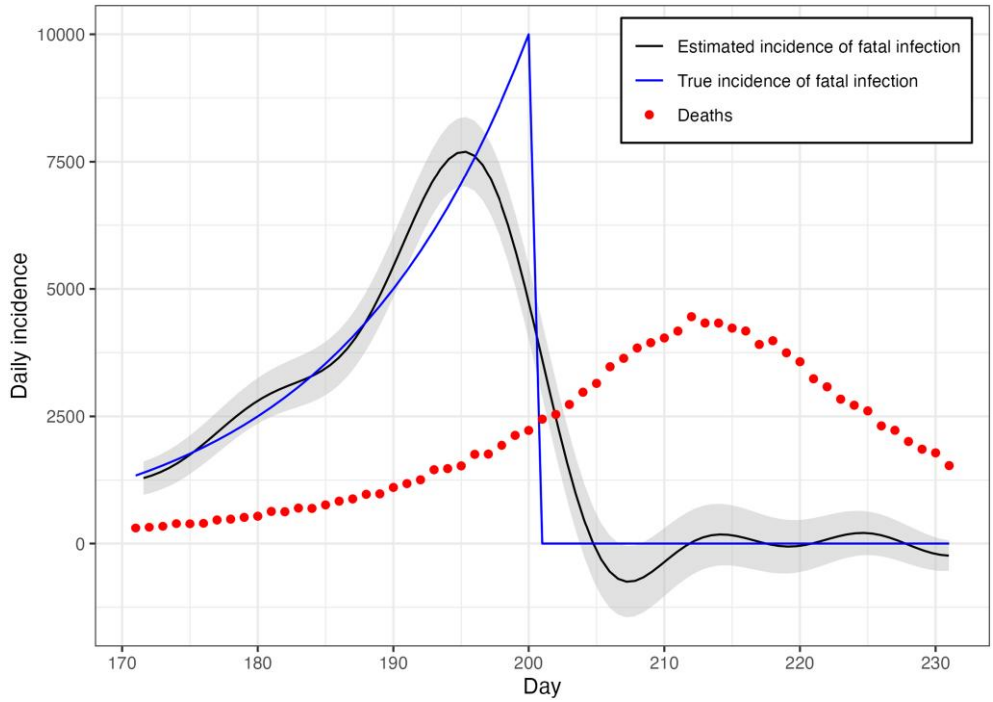



**Figure 1.** Estimated incidence of fatal infections using the deconvolution framework of Wood et al. under a contrived simulation of an abrupt crash to zero incidence. Using parameter settings matching their original specification (`k = 300, lmu = 3.152, lsd = 0.451`), the true peak is day 200 but the estimated peak is shifted backwards to day 195.

demands on the scientific community. The paper by Wood et al. highlights several critical issues that emerged from this time, including media distortions, the framing of risk and uncertainty, and the need for a critical appraisal of the assumptions underpinning scientific evidence. We strongly agree that reflecting on these challenges is essential for strengthening preparedness for future public health threats.

However, we contend that the paper by Wood et al. overlooks several important considerations and introduces assumptions of its own. In particular, we question whether the deconvolution framework, while parsimonious, is sufficiently flexible to address all questions regarding the precise timing of infection peaks—and by extension, the impact of interventions such as lockdowns. Just as the approach of Flaxman et al. imposed an assumption of abrupt changes in incidence, the approach of Wood et al. imposes an assumption of smoothness, which inevitably shapes the resulting conclusions. This effect is illustrated in Figure 1, which presents a hypothetical simulation of an abrupt crash in infections to zero. Using the code and parameters provided by Wood et al., the method underestimates the timing of peak incidence by approximately five days.

While the assumption of smoothness can certainly be relaxed—through, for example, a Gaussian process with an exponential kernel or a discretized random walk process (Bhatt et al., 2023)—our point is not that the problem is unsolvable, nor that this reflects the true pattern of infections in the UK during lockdowns. Rather, we argue that the conclusion of an earlier infection peak by Wood et al. is, at least in part, a product of their modelling assumptions. The assumption of perfect smoothness is demonstrably flawed, just as the assumption of an abrupt change is flawed; the reality likely lies somewhere between these extremes. The crucial question, therefore, is whether the available data contain sufficient signal to reliably distinguish between these competing modelling choices.

This leads us to a second concern with the deconvolution approach: its reliance on a single data stream—daily deaths. While this may have been a reasonable strategy in the early days of the pandemic, when data were limited and models had to be developed under intense time pressure, it is

less defensible 5 years after the fact. Important questions such as those raised by Wood et al. warrant answers that draw on a broader range of data sources, including those that are more directly informative about transmission dynamics. For example, if the hypothesis is that infections peaked prior to the lockdown announcement due to voluntary or anticipatory behavioural changes, then it should be evaluated using additional evidence—such as mobility data, seroprevalence studies, healthcare service utilization data, or other innovative data sources. Relying solely on daily deaths within a single, unstratified analysis stretches this data stream beyond the limits of what it can reliably inform.

More broadly, while the current method identifies a single surprising and interesting result—an earlier-than-expected peak in infections—which emerges in part from the modelling assumptions themselves, what is missing is a positive, mechanistic explanation for the underlying dynamics. In the absence of such an explanation, it is difficult to assess the plausibility or generalizability of the conclusions. Strong claims require not only identifying anomalies but also articulating coherent alternative hypotheses that are consistent with a broader body of evidence. Without this, the argument risks becoming a post hoc rationalization of an artefact of model structure, rather than a robust inference about real-world epidemic processes.

We thank Wood et al. for their contribution and believe that their important questions will ultimately strengthen infectious disease modelling methods.

*Conflicts of interest:* None declared.

## References

Bhatt S., Ferguson N., Flaxman S., Gandy A., Mishra S., & Scott J. A (2023). Semi-mechanistic Bayesian modelling of COVID-19 with renewal processes. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *186*(4), 601–615. https://doi.org/10.1093/jrsssa/qnad030

Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., Monod M., Perez-Guzman P. N., Schmit N., Cilloni L., Ainslie K. E. C., Baguelin M., Boonyasiri A., Boyd O., Cattarino L., … Bhatt S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

https://doi.org/10.1093/jrsssa/qnaf108
Advance access publication 2 July 2025

# Michael Whitehouse, Lorenzo Rimella, and Nick Whiteley's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Michael Whitehouse[1], Nick Whiteley[2] and Lorenzo Rimella[3]**

[1]School of Public Health, Imperial College London, London W12 0BZ, UK
[2]School of Mathematics, University of Bristol, Bristol BS8 1UG, UK
[3]ESOMAS, University of Torino 10134 and Collegio Carlo Alberto 24127, Torino, Italy

*Address for correspondence*: Michael Whitehouse, School of Public Health, Imperial College London, London, UK. Email: michael.whitehouse@imperial.ac.uk

We commend the authors of Wood et al. (2025) for their thorough contribution to the retrospective analysis of the statistical response to the Covid-19 crisis. We focus on the issues raised in Section 4 around the use of models with increasing extents of heterogeneity. The authors correctly recognize that omitting such considerations can lead to incorrect inference, e.g. poor estimation of final size. It is important, however, to acknowledge the challenges associated with learning model parameters from available data.

As opposed to aggregated incidence/prevalence counts in classical epidemic models and metapopulation models (Danon et al., 2021; Kermack & McKendrick, 1927), individual-based models (IBMs), such as the influential CovidSim model (Ferguson et al., 2020), explicitly consider individuals' disease states. This provides a refined representation of the population and in principle allows practitioners to work with covariates defined at an individual level (e.g. health records, geographical locations, contact networks). A simple example is a regression modelling of the rates at which the $k$th individual infects the $n$th: $\log \beta_{nk} = \log \beta + \mathbf{c}_n^\top \mathbf{b}_S + \mathbf{c}_k^\top \mathbf{b}_I$, where $\mathbf{c}_n$, $\mathbf{c}_k$ are observed covariates associated with the $n$, $k$th individuals and $\beta$, $\mathbf{b}_S$, $\mathbf{b}_I$ are model parameters respectively controlling background infection and how covariates contribute to attracting (susceptibility) and transmitting (infectivity) infection. This sliding scale of heterogeneity can be linked to the population distribution of the individual specific susceptibility parameter $\alpha$ of the model defined in subsection 4.1 of Wood et al. (2025).

Since transmission processes are typically only partially observed, exact statistical inference, e.g. aimed at learning regression parameters $\beta$, $\mathbf{b}_S$, $\mathbf{b}_I$ or informing properties of $\alpha$, requires marginalization over the latent disease state of the entire population. This incurs a prohibitively expensive cost which is exacerbated by increasing individual heterogeneity, motivating the development of approximate inference methods (Bu et al., 2024; Ionides et al., 2023; Rimella, Jewell, et al., 2025; Rimella, Whiteley, et al., 2025; Whitehouse, 2025; Whitehouse et al., 2023). Furthermore, the identifiability of models incorporating individual heterogeneity, in general, depends on the observation model, for instance whether individual test results or aggregated incidence/prevalence counts are reported. In particular, Whitehouse et al. (2023) and Rimella, Whiteley, et al. (2025) show that, in the context of both metapopulation models and IBMs, parameters defining processes with coinciding large population behaviour cannot be distinguished.

We are curious about the balance between realism and feasibility and how the research community will address this in the future: given the prohibitive cost of fitting increasingly heterogeneous models, how far should modellers go? In the presence of identifiability issues, how should modellers acknowledge and account for these limitations? Will further development of efficient inference methods, identifiability conditions, and optimal observation sampling, e.g. in a Bayesian experimental design framework, benefit future outbreak response efforts?

*Conflict of interests:* None declared.

## References

Bu F., Aiello A. E., Volfovsky A., & Xu J. (2024). Stochastic EM algorithm for partially observed stochastic epidemics with individual heterogeneity. *Biostatistics*, *26*(1), kxae018. https://doi.org/10.1093/biostatistics/kxae018

Danon L., Brooks-Pollock E., Bailey M., & Keeling M. (2021). A spatial model of COVID-19 transmission in England and Wales: Early spread, peak timing and the impact of seasonality. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200272. https://doi.org/10.1098/rstb.2020.0272

Ferguson N. M., Laydon D., Nedjati-Gilani G., Imai N., Ainslie K., Baguelin M., Bhatia S., Boonyasiri A., Cucunubá Z., Cuomo-Dannenburg G., Dighe A., Dorigatti I., Fu H., Gaythorpe K., Green W., Hamlet A., Hinsley W., Okell L. C., van Elsland S., …Ghani A. C. (2020). *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*, Vol. 16. Imperial College London, London.

Ionides E. L., Asfaw K., Park J., & King A. A. (2023). Bagged filters for partially observed interacting systems. *Journal of the American Statistical Association*, *118*(542), 1078–1089. https://doi.org/10.1080/01621459.2021.1974867

Kermack W. O., & McKendrick A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *115*(772), 700–721.

Rimella L., Jewell C., & Fearnhead P. (2025). Simulation based composite likelihood. *Statistics and Computing*, *35*(3), 58. https://doi.org/10.1007/s11222-025-10584-z

Rimella L., Whiteley N., Jewell C., Fearnhead P., & Whitehouse M. (2025). 'Scalable calibration for partially observed individual-based epidemic models through categorical approximations', arXiv, arXiv:2501.03950, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2501.03950

Whitehouse M. (2025). 'Accelerated inference for stochastic compartmental models with over-dispersed partial observations', arXiv, arXiv:2505.06935, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2505.06935

Whitehouse M., Whiteley N., & Rimella L. (2023). Consistent and fast inference in compartmental models of epidemics using poisson approximate likelihoods. *Journal of the Royal Statistical Society: Series B, Statistical Methodology*, *85*(4), 1173–1203. https://doi.org/10.1093/jrsssb/qkad065

Wood S. N., Wit E. C., McKeigue P. M., Hu D., Flood B., Corcoran L., & Jawad T. A. (2026). Some statistical aspects of the Covid-19 response. *Journal of the Royal Statistical Society: Series A*, *189*(1), 1–31. https://doi.org/10.1093/jrsssa/qnaf092

https://doi.org/10.1093/jrsssa/qnaf097
Advance access publication 1 July 2025

# Jason Wyse, Dhorasso Temfack, Eishita Yadav and James Sweeney's contribution to the Discussion of 'Some statistical aspects of the Covid-19 response' by Wood et al.

**Jason Wyse[1], Dhorasso Temfack[1], Eishita Yadav[2] and James Sweeney[2]**

[1]School of Computer Science and Statistics, Trinity College Dublin, D02 PN40 Dublin, Ireland
[2]Department of Mathematics and Statistics, University of Limerick, V94 T9PX Limerick, Ireland

*Address for correspondence*: Jason Wyse, School of Computer Science and Statistics, Trinity College Dublin, D02 PN40 Dublin, Ireland. Email: jason.wyse@tcd.ie

We congratulate the authors on a thought-provoking and timely paper evaluating the effectiveness and impact of decisions made during the COVID-19 pandemic on society at large. The authors produce compelling arguments as to why some of the approaches taken to modelling, model checking, and prediction may not have been optimal. They suggest that narratives around disease risks and the benefits of disease mitigation strategies, communicated to the public, did not meet statistical best practice. We fully agree with their sentiments on the inappropriateness of data being made available only to siloed groups as opposed to the research community at large.

However, while robust, we do wonder if the authors' criticisms of the models that were used fully appreciate the context in which they were built and deployed. We have applied the Section 5 code to Irish death case data where it performs well on a long time series of 1.5 years, matching the timeframe of the analysis in the paper. However, at pandemic outset where data is sparse and the pathogen poorly understood, how well would the approach work from an advisory perspective? The first lockdown was announced in the UK in March 2020; however, we note that the paper is reliant on a metanalysis (McAloon et al., 2020) that was published in July 2020. How would the authors' approaches have performed using contemporaneous estimates of infection to death duration and data available at time of first lockdown in March 2020? We also wonder how the authors' framework would perform for diseases characterized by abrupt changes in the disease dynamics, especially to reconstruct the change in $R_t$ (Section 5 of the paper). Can the smoothness assumption of the model lead to underestimating the sharpness of real changes, causing the

estimated $R_t$ change to appear as though it starts earlier and ends later than may actually be the case? There is, however, considerable food for thought in the paper. It is undisputed that countries worldwide were not as prepared as they may have been. Given the societal upheaval of COVID-19, preparation for the next pandemic is vital towards ensuring economic and social stability.

Despite the paper offering many interesting critiques and discussion points, we felt there was some missed opportunity in providing concrete guidance on what might be done differently when such a crisis scenario arises again. In such circumstances, could retrospective analysis like that reported in Section 5 provide useful insights? For example, if a novel pathogen emerges for which a vaccine is not available, what is statistical best practice to inform the public and policy makers on foot of the authors' analysis? How should the population be sampled to estimate case incidence? How should key transmission settings be identified in real-time?

*Conflicts of interest:* None declared.

## Reference

McAloon C., Collins Á. Hunt K., Barber A., Byrne A. W., Butler F., Casey M., Griffin J., Lane E., McEvoy D., Wall P., Green M., O'Grady L., & More S. J. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open*, *10*(8), e039652. https://bmjopen.bmj.com/content/10/8/e039652. https://doi.org/10.1136/bmjopen-2020-039652

**The author replied later in writing as follows:**

https://doi.org/10.1093/jrsssa/qnaf110
Advance access publication 2 July 2025

# Authors' reply to the Discussion of 'Some statistical aspects of the Covid-19 response'

**Simon N. Wood[1], Ernst C. Wit[2] and Paul M. McKeigue[3]**

[1]School of Mathematics, University of Edinburgh, Edinburgh, UK
[2]Institute of Computing, Univerità della Svizzera italiana, Lugano, Switzerland
[3]College of Medicine and Veterinary Medicine, University of Edinburgh, Edinburgh, UK

*Address for correspondence*: Simon N Wood, School of Mathematics, University of Edinburgh, Edinburgh EH9 3FD, UK. Email: simon.wood@ed.ac.uk

## 1 General response

*…humans find it very difficult to consciously reflect on a large number of datapoints and weigh them against each other …That's why when faced by complex issues—whether a loan request, a pandemic or a war—we often seek a single reason to take a particular course of action and ignore all other considerations. This is the fallacy of the single cause.* Yuval Noah Hari *Nexus*.

What went well in the UK Covid response? This question, posed by Jose and others, is a good place to start, given our paper's explicit focus on what did *not* go so well. Here are some of what we believe are scientific highlights.[1] The ONS survey, *measuring* the state of the epidemic was internationally leading: why it was not widely replicated internationally is a mystery. Similarly REACT and REACT2 and the openness of much other data were exemplary. The central UK government not pursuing the pseudoscience of zero Covid and, perhaps through growing model

scepticism, declining to lock down for Omicron were also huge positives. The UK's lab based biomedical scientists, like those of several other countries, obviously did a great job on vaccines.

We would like to thank all the people who have taken the time to comment on a paper that for many would have been much more comfortable to ignore—although it is a pity that Sir Chris Whitty, Dame Angela McLean, Sir Ian Diamond, Neil Ferguson OBE and Grahame Medley OBE were too busy (or otherwise unwilling) to contribute when invited to do so. Those closest to the decision processes will surely have had the most coherent counter arguments to what we have written and it would have been good for those to be public. Also absent from the discussion are the trenchant criticisms of the UK Covid response that a number of respected statisticians are willing to make in private. We think that it is time to make those public, although we can well appreciate the reluctance to do so.

In responding, we begin by noting that none of the critical comments directly addresses the central issue of our paper: that every Covid intervention involved not only benefits, but also significant costs. Based on preexisting data and prior research, it should have been apparent, early in the pandemic, that the long-term health (and other) risks stemming from the economic consequences of the measures were comparable to—or even greater than—the health risks those measures aimed to mitigate.

Effectively ignoring these trade-offs and endorsing a maximal remediation strategy required an appeal to an absolute imperative: in the UK, avoiding the collapse of the NHS.[2] The unacceptability of that scenario apparently justified virtually any future cost. However, such an absolutist rationale is only defensible if the maximal measures were indeed *necessary* to prevent the outcome, and the most reliable reconstructions of incidence suggest that this was probably not the case. Without this, the justification for risk-distorting fear-messaging, the suppression of dissent as misinformation, the up-ending of usual evidence norms in science, the suspension of civil liberties and the damage done to the young in the interests of the old, are even more difficult to view as sound, evidence based, public health measures.

Where did this approach come from, of near zero risk appetite in the short term combined with enormous risk appetite over longer time frames? The latter, at least, is somewhat consistent with the response to climate change. It is scientifically clear that exacerbating climate change will come with truly existential medium to long-term risks. Nevertheless, whereas the UK Covid public spending alone of £310–410 Billion has received rather limited scrutiny, a planned UK spending package of £27 Billion, to tackle climate change, was canceled in 2024 as now *unaffordable*. In this context, it is also concerning that poor scientific messaging and modelling around Covid has fueled somewhat justifiable public mistrust, including a far less justifiable scepticism about climate messaging and modelling.

Moving on to the many detailed comments raised, the most widespread, legitimate and substantive technical criticism is that we failed to allow for improvements in life expectancy in the iterated life table method for calculating excess deaths, when the steady rise seen pre-Covid made this a reasonable assumption (McDonald, Fisch, Dodd et al., Andrews and a number of in person comments after the meeting). This point is completely fair, and very easy to correct using the life table method and supplied code, by simply reducing $d_w$ each year by the factor required to increase life expectancies by the expected amount. Using the pre-Covid figure of 18 days life expectancy increase per year, results in around 28,000 more cumulative excess deaths than we calculated in the paper, as can very easily be checked using our supplementary code. The objection was also raised that we accumulated excess deaths from the start of 2020. Given Sheila Bird's point, that the first known UK Covid death eventually turned out to have been on 31 January 2020, we are not sure that this is unreasonable.[3] Nevertheless, if we instead choose to start accumulating from the excess death minimum in 2020, then nearly 6,000 excess deaths are added to the total, taking us to around 130 thousand. The discrepancy with methods that do not fully account for an ageing population (or life expectancy improvement) is still very large: our Figure 5 shows why this has to be the case.

A second important objection to the work comes from Paul Birrell, Lorenzo Pellis, Robert Verity and their colleagues. That is the concern that the infection peak timing results could be smoothing

[2] What this meant was not spelled out—perhaps being forced into a mode of operation normally seen only in war zones?

[3] The criticism of quoting the later date when discussing ONS first sampling dates seems less reasonable: no one in April 2020 could be blamed for responding too slowly to a death not discovered until September 2020.

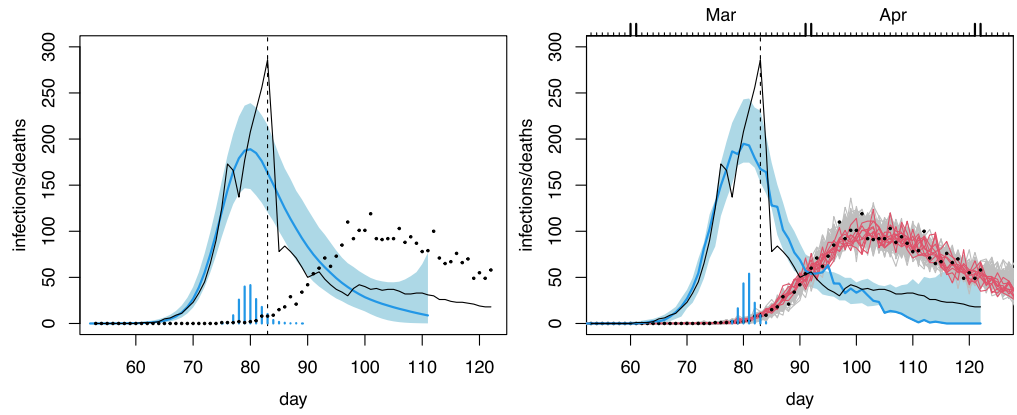



**Figure 1.** Left: Replication of the Birrell et al. simulation shown in their lower panel Figure 2, using their assumed incidence profile, but also showing the uncertainty in peak location for the Wood (2021) reconstruction shown. The blue bar chart shows the posterior distribution of the peak location (scaled), which clearly includes the truth. Right: equivalent reconstruction using the simulation based deconvolution method of Wood (2025), which does not make a smoothness assumption (red trajectories are replicate death trajectories simulated from the dark blue incidence reconstruction, grey are bootstrap replicates of the same). Again there is insufficient information in the signal to resolve the high frequency feature of the simulated incidence, but in this case the misstiming is reduced, while the distribution of inferred peak time again includes the truth.

artefacts if the real incidence was in fact exponential increase until lockdown followed by collapse. Wood (2021) examined this question in detail (as did the paper's arXiv version from June 2020), covering exactly the point Birrell, Verity et al. discuss and demonstrating that the effect is insufficient to explain the results on the real data. Wood (2025) also provides a method that eliminates the smoothness assumption and provides a statistical test for hypothetical incidence profiles: the timing results do not change and, for England and Wales, a profile of exponential incidence growth with collapse at lockdown is firmly rejected. In any case, such an inference artefact cannot apply to the direct sampling based estimates from REACT2 and the ONS (or the direct estimates from Wuhan, in Figure 2, showing a similar pattern).

Several discussants also provide simulations intended to support their point. The Pellis simulations do not really seem relevant. Figure 1 of his study shows the fit of an obviously incorrect smooth model, from which incorrect $R$ estimates are then derived: what else is to be expected? Furthermore, any such misfit can easily be identified from appropriate residual plots. The peak displacements seen in the Verity et al. simulation are consistent with the discussion in Wood (2021), given the slower epidemic dynamics that they consider: their simulated doubling time of 10 days is much longer than that which pertained in the first Covid wave. So, the discussion of this issue in the context of the real data remains unaffected. For one simulation, Birrell gets an incidence reconstruction mistimed by 4 days, based on a truth that increases rapidly, then decreases sharply for 2 days and then increases again for the week until lockdown. Any method will have difficulty resolving such a feature from data obtained by convolving it with the time to death distribution,[4] but what Birrell has not shown is the uncertainty of the estimated peak location, which is then high, partly also because of the low numbers in the simulations. Figure 1 corrects this omission, while its right panel shows the result of applying (Wood, 2025) to the same data. Birrell's regional point echoes (Wood & Wit, 2021), but the uncertainty his simulations reveal at lower levels of aggregation emphasize the difficulty of separating what is data driven and what model driven from more local data. Actually, it is exactly the intrinsic uncertainty associated with relatively low counts that makes a daily deaths deconvolution for Wuhan too uncertain to be useful, but it is worth noting that Wuhan actually provides another data source for reliable direct incidence reconstruction: first symptom onset date was recorded for all fatal cases.[5] Figure 2 deconvolves these data with the

[4] For example, how would the MRC method fare, given its assumption of weekly step changes with smoothness imposed by a random walk prior?

[5] We are grateful to Piet Streicher for pointing this out to us after the meeting.

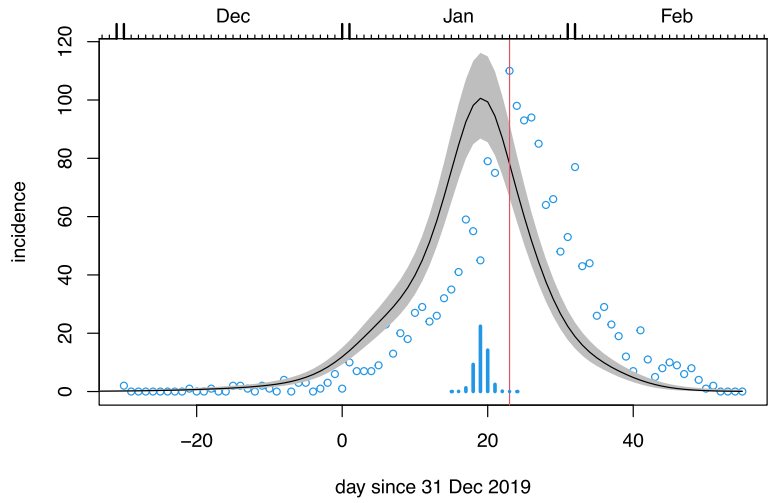



**Figure 2.** Wuhan fatal incidence reconstructed by deconvolution (black and grey) from the recorded time of first symptom onset for fatal Covid cases. Blue circles are the number of fatal cases with onset on the given day (China CDC data, Bai et al., 2020). The red line is at the date of Wuhan lockdown on 23 January. The blue bar chart is the simulated posterior distribution of time of peak incidence (scaled) corresponding to a 95% CI of 17–21 January 2020.

infection to onset distribution to reconstruct incidence (the much narrower distribution greatly eases the deconvolution task).

Fundamentally, the problem with smoothing is not the smoothness *assumption* (which can anyway be removed as in Wood, 2025), but the smoothing *operation* inherent in the convolution of incidence with the disease duration distribution. The difficulty of reversing that is what makes it difficult to distinguish the fit of a discontinuous incidence curve from a smooth one (but again see Wood 2025 for a statistical test). Once that is recognized, there is no justification for asserting that model fitting strongly supports the narrative of surging incidence up until lockdown. It cannot, and as soon as it was demonstrated (in early May 2020) that a continuous incidence curve peaking well before lockdown explains the data at least as well, there was no excuse for continuing with the assertion. Further, the earlier peak model was consistent with the data that was then available on the extent of contact rate reduction, the drop seen in NHS 111 calls for respiratory diseases and the fact that Sweden's wave had evidently turned within a day or two of the UK, without lockdown. The careful checking of the possibility of a smoothing artefact in Wood (2021) also suggested that the peak before lockdown result was likely to be real.

In our view, from the moment in May 2020 that the simple deconvolution results first appeared it was not reasonable or responsible to continue arguing that the evidence strongly supported lockdown having caused incidence to decline. It was at least as plausible (more likely given other data) that incidence was in sharp decline before lockdown. Given the evidence that has accumulated since then, is it really reasonable to insist that the question is still unresolved?

The preceding sentence benefits from hindsight, of course, something that several discussants mention about the paper more generally. We actually think that much of what we discuss suggests poor risk management *given what was known at the time*, but of course some things, like the depth of the economic damage, or that 'build back better' was an illusion, or the damage to school aged children, are now much clearer than they were in 2020. However, we think that the most significant piece of true hindsight, framing current discussions, is the fact that it proved possible to produce the first vaccines against a corona virus and to do so in a very short space of time.[6] At the pandemic outset, it is difficult to see a rational basis for assuming that such a development was more likely than not, let alone certain. Given that significant uncertainty, Sweden's approach of managing Covid as a part of everyday life appears rational, but the UK's approach less so.

[6] Albeit the vaccines could not produce the herd immunity that was assumed in the discussions of 2020.

We now turn to the three invited discussants, Goldblatt, DeAngelis and Bird[7] : Peter Goldblatt's most substantive point seeks to disconnect the life loss identified in the Marmot report from the economic disruption caused by the 2008 crisis. Fundamental to this point is a belief that austerity policies were unrelated to the financial crisis. We do not find this credible, but readers can decide for themselves. We respond to his other points in the detailed response section.

Daniela De Angelis' contribution (with Kandiah and Birrell) makes some points with which we agree, but expresses one opinion with which we do not: that careful model choice, validation and quantification of uncertainty were luxuries that could not be afforded in a crisis. How, in that case, can one avoid the sort of 'reckless quantification' that fails the most basic requirements to 'first do no harm?' If there is not sufficient validation to know whether a model is remotely suitable for prediction, or whether the reported uncertainties are plausibly well calibrated, then what is a model but a mathematically stated opinion? Or worse, an opinion gaining spurious credibility from the superficial appearance of statistical validity? As in the 2008 financial crisis, we think that managing risk with models whose reliability is oversold has the potential to do great damage, not least because the lure of apparent precision then leads to the down-weighting of other evidence and other factors for which the uncertainties have been more honestly stated.

De Angelis' contribution also makes one substantial technical point that at first sight appears to call into question our results on timing of the incidence peaks (and $R < 1$) relative to lockdown. They report results from Kandiah et al. (2025) which update Birrell et al. (2021) by allowing contact rates to change weekly before as well as after lockdown, thereby removing the feature of Birrell et al. (2021) that surging incidence up until lockdown was simply built into the model. They report that they still estimate $R$ to be substantially above 1 and incidence to be soaring until lockdown, but they also report results consistent with our paper if they use the same infection to death distribution.[8] This begs the question of what infection to death distribution Birrell et al. (2021) and Kandiah et al. (2025) used? Neither paper or the accompanying supplementary material tell us, simply referring to '*an assumed-known distribution of the time from infection to death from COVID-19* f' in the Likelihood sections of the supplementary materials. No reference is given, although later in the Birrell et al. (2021) supplementary material it is reported that the mean incubation period was assumed to be 4 days (s.d. 1.41) and the mean time from symptom onset to death was assumed to be 15 days (s.d. 12.1). The papers' cited code repository confirms these numbers and indicates that a gamma distribution was used for the latter distribution, although the Birrell et al. (2021) code also has an option for a mean 17.8 and s.d. 8.9. Kandiah et al. (2025) do not appear to state what the distribution was in the paper or supplementary material, but the Kandiah et al. (2025) code uses a gamma distribution with options of the Birrell parameters (mean 15 with s.d. 12.1), or of mean 9.3 with s.d. 9.7, or of mean 9.0 with s.d. 8.1 (this last option is commented '*Currently highly unprincipled*').

Given the lack of information on where these distributions came from, all we can do is plot them alongside the distributions with known sources given in our paper. Figure 3 shows that they are inconsistent with the distributions from published sources, and the high numbers dying a day or two after symptom onset seem biologically implausible. Additionally, if these distributions were considered reasonable then there is also an awkward mismatch with the incidence estimates based on direct sampling to explain, which does not occur if we use the distributions from published sources. The assumed incubation period distribution also appears shorter than the literature suggests.

The models used in Kandiah et al. (2025) perhaps lean towards a type more common in mathematical biology than statistics. Some important details are parameterized in an ad hoc manner from disparate data sources, with some strong assumptions having to be made. Comparatively few parameters are inferred statistically and it is unclear how that inference is then influenced by the other assumptions and estimates. In particular, key contact processes are *assumed* to change at lockdown, apart from those that are estimated. Hence it is unclear to us that the model is really suitable for reliably estimating changes in incidence or $R$ around lockdown, even with more

[7] The meeting did not follow the Society's usual proposer-seconder tradition.

[8] Figure 3 $R$ results of their study are consistent with ours, given the averaging over the week preceding lockdown that their weekly step function entails; Figure 4 of their study is consistent with the regional results in Wood and Wit (2021) aggregated in our Figure 14. The correspondence would presumably have been clearer if incidence had also been plotted for these scenarios. Note that the northwest is the last region for which $R$ drops below 1.

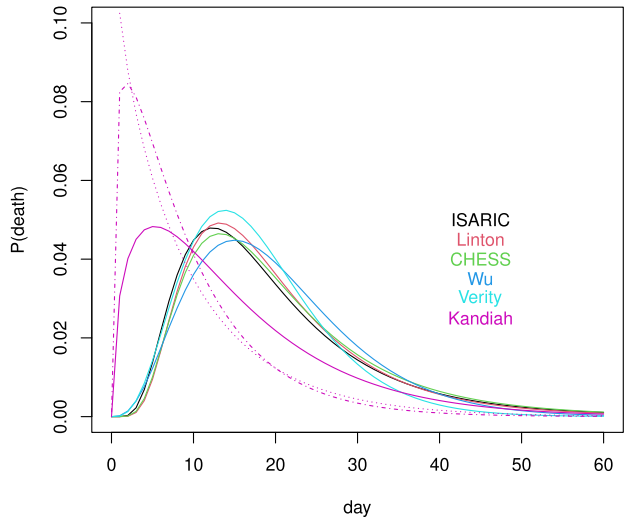



**Figure 3.** As Figure 9 in the main paper, but with the symptom onset to death distributions from the Kandiah et al. (2025) code shown in pink. The solid pink curve is also the distribution used by Birrell et al. (2021). No references or other source information appear to be given for the pink distributions. Note that the unused distribution option in the Birrell code with mean 17.8 would be completely consistent with the plotted distributions from known sources.

reasonable disease duration distributions. It does not seem suitable for producing the claimed *counterfactuals*. Given limitations of space here we would urge the reader to read Kandiah et al. (2025) and Birrell et al. (2021) and form their own view.

Daniela De Angelis also invited Kevin Fong to present at the meeting. He provided a masterful demonstration of the power of hard hitting emotional messaging over dispassionate consideration of risk, leaving us with the following questions: are there *any* limits to the future cost in lives and livelihoods that would be justified to mitigate the trauma to health service staff that a pandemic inflicts? Secondly, does employing emotionally charged messaging to challenge quantitative risk assessments align with the imperative to avoid causing unintended harm? His written comments are more measured, underlining that the opening statement of our paper is not hyperbole.[9] What he does not do is to provide any actual evidence that the measures taken were *necessary*, nor that there was an appropriate balance between 'protect the NHS' and the limitation of other societal damage, including to the development and education of school age children, the future health and wellbeing of the less economically advantaged and the long-term economic activity that in the end funds the NHS.

Many of Sheila Bird's points we agree with or are fair comment, but there was also a point about the appropriate multiplier on the NICE threshold and what is appropriate during war time. We are not sure that the exact multiplier really changes the point that a large discrepancy may suggest something suboptimal. Further, it is the fact that war is societally *existential* that justifies the suspension of all usual norms. Presumably no one would still argue that Covid was that, but what if a disease did emerge that was an existential threat? That is both highly contagious and with an infection fatality rate to all ages in the 25–50% range that is not historically impossible. The NICE threshold would then suggest spending 10–20 times GDP, which is obviously not remotely realistic (it comfortably exceeds total national wealth, for example). The war time analogy raises another awkward question. A contested full scale invasion of the UK would presumably produce casualties at a level that would overwhelm the NHS. Does that mean that the only acceptable option would be immediate capitulation? We agree that of course a pandemic demands exceptional outlay, but that does not remove the need for proportionality. Will it end up being judged proportionate to

[9] As context to some of his figures: the general mortality rates after ICU admission are around 30%, with 40% being about the rate seen for intracerebral haemorrhage ICU patients (van Valburg et al., 2020); health workers suffered Covid death rates around the median of all ONS occupational categories, but this constituted an increase on background death rates by occupation that was (just) at the top of the distribution across the categories.

have taken an economic hit of the order of £$10^{12}$ combatting Covid, and to have then decided, in the light of the consequent debt burden, that planned spending of around 3% of that, to combat the threat of climate change, was no longer affordable?

Moving on to some contributions more aligned with the paper's message, Ioannidis, Di Loro et al., Berger et al., and others all discuss the difficulties in getting solid, evidence based narratives heard in the media and elsewhere during the pandemic, with scientific journals also failing to offer the balance and rigour that one might hope. Ioannidis argues that picking apart how this came about will require a combination of statistics, psychology and sociology. We think he is right. For example, the element of (social) media driven social contagion in the wave of panic over Covid, that swept the developed world and beyond in March 2020, seems at least as significant as the actual risk, serious though that was. Some of the reason for our susceptibility to such contagion must lie in the obvious adaptive advantages of believing what many other people believe, combined perhaps with a tendency to judge the reliability of others on the basis of their adherence to views we view as sensible. One such view is 'the wisdom of crowds', often supported by Galton's observation that the median of 800 county fair competitor's guesses for the weight of an ox was within 1% of the true weight. But we suspect that an experienced crowd, making an educated guess of a simple quantity, may not be a good model for a social media crowd pronouncing on a complex topic about which they know little. It is likely true that someone in the crowd will have sensible answers, but equally unlikely that they will get heard above the noise. And that noise was often loudest from the large numbers of suddenly emerging experts, often desperate for the media exposure that might count as academic career enhancing 'impact'. In this context, a retired GP remarked that he felt that Covid had unleashed an epidemic of the Dunning–Kruger effect.

Similarly significant is the human desire for control, and the tendency to see events as the result of our actions, even when they are not. One could view the continued insistence on the key role of lockdowns in this light, but in the UK context even the general view of the role of science in the initial lockdown decision perhaps suffers from this illusion of control. At least if Dominic Cummings' evidence to the Covid Inquiry is to be believed (Cummings, 2023), the decision to lock down was not primarily driven by the UK government's official scientific advice (such as Imperial college report 9), but by Cummings and an informal group of four 'incredibly able people' he had turned to. The advisors he lists are indeed outstanding in their respective fields. But those fields do not include epidemiology, public health, statistics or risk management, while, however well read, Cummings' training was in ancient history rather than science. As Cummings tells it, he was particularly influenced by the advice of a pure mathematician, working with an exponential growth model for Covid, who urged him to 'push extreme suppression immediately' on 14 March 2020. Cummings then apparently repeatedly warned the prime-minister of the risk of a 'zombie apocalypse', and on 19 March that, without lockdown, 'the NHS in London collapses in 15 days' (c.f. Figures 10–14 in our paper).

The pressures and incentives operating on scientists and decision makers also need serious thought. There is the obvious pressure to make predictions that are broadly correct. The prediction that maximum suppression of human contact will result in suppression of viral transmission is safe in this regard. In contrast any prediction about how little suppression would be sufficient is much harder to get right. The asymmetry in blame is also rather extreme: don't do enough and you will for sure be seriously blamed. Do too much and it is always possible to argue that it wasn't. Also, scientists are human, operating in the same media, social-media and social environments as everyone else. The pressure to do work supportive of what is believed to be right by right thinking people is very high, and, especially in the age of social media, the pressure not to publish contrary work is similarly strong. And that is before we get to the baleful effects of the promotion process on academic freedom of thought.

To end on a constructive note, and in response to Keeling, Wyse and Srakar's questions about lessons to be learned for the next pandemic, we have some suggestions.

1. While the Imperial group and others clearly swung into action early, generally the government led scientific response was sluggish, leading to the panic at the heart of government described in Cummings' evidence to the UK Covid inquiry. The obvious lesson for next time is

the need to ramp up activity much earlier, so that policy can be based on science, rather than the ad hoc approach described by Cummings.
2. Do not base quantitative management on un-calibrated models that were not developed for prediction, but rather as very useful aids for structured reasoning and gaining qualitative insight (see e.g. Streicher et al., 2025).
3. Echoing Britton, Irons and Maruotti et al., risk management should not become focussed on a single or small set of objectives simply because competing risks are much harder to quantify with the accuracy claimed for the epidemic itself.
4. Rather than use psychology for manipulating emotions to undermine risk assessment, it would be preferable to focus on the many things that psychology tells us about the confirmation, selection and other biases, to which we are all prey in assessing data and risk, and the way that these can interact with group dynamics to make it difficult to re-assess decisions, once made. In particular, it would probably be a good idea to have an independent advisory group complementary to SAGE/SPI-M whose *remit* is to question the science and to avoid focussing on too narrow or short term a set of societal objectives.
5. The role of statistics is not to get better parameter estimates to feed into 'the models', it is to properly assess (and make decisions given) uncertainty and risk, and provide the methods by which the hypotheses and assumptions encoded in models can be rigorously empirically tested against data.
6. Statistical methods need to be properly understood to provide useful insight—fitting a model to data is only part of statistical inference, and, as Maruotti et al. emphasize, good fit does not mean good predictive validation. Particularly important are the sampling and design theory necessary for understanding when a set of data form a valid basis for answering a scientific question of interest. Without this, statistical methods can be quite dangerous. For example, likelihood is not a piece of magic to allow models to be fit to data, but fundamentally a model for the random sampling process by which the data are to be treated as being produced.
7. Measuring a problem is usually better than modelling it: to this end the UK should have a preapproved surveillance survey and cohort ready to go, with protocols for batch testing thought through for the scenario in which test capacity is limited.
8. It would be good to fill some of the data gaps well before the next pandemic. For example, the impressive POLYMOD survey (Mossong et al., 2008) used to parameterize age-specific contact rates in many epidemic models was based on 1-day diaries kept by around 1,000 UK residents of whom 7 were over 75 years and none over 80 years of age. Such a data deficit should probably be a much higher priority for rectification than method development.
9. A situation serious enough to justify the closing down of society should surely justify the relaxation of some of the more stringent aspects of data privacy rules to ensure that policy relevant data are open and available for the common good.
10. Try to avoid the yearning for certainty compromising the honest assessment of uncertainty.

We hope that, despite its necessarily somewhat negative tone, our paper can contribute to the discussion about what the most important lessons were, and that when the next pandemic arrives at least some of David Spiegelhalter's 'many regrets' can be avoided.

## 2 Response to individual contributions

Here we provide brief responses (in alphabetic order of first author last name) to all the discussants, except where the main points are already covered above.

**Allorant** argues that we only focussed on mortality, ignoring morbidity. Although this is a fair point, we note that we also did not consider morbidity effects of the NPI measures, which are unlikely to have been minor given the tendency for economic deprivation to shorten *healthy* lifespan more than lifespan as such. The argument that mortality maybe underestimated based on Msemburi et al. (2023) is extraordinarily weak. See Figure 6 of Msemburi, which makes it clear where they believe undercounting occurred and that it is certainly not the UK. Even ignoring all the other evidence for tiny IFR/CFR in the young, what is Allorant's explanation for the low observed deaths in the young if they had high CFR? Were they simply not catching Covid (in which case why lock them down)? Given the unused Nightingale hospitals and the evidence that infections were on the way down before lockdowns we are unsure that the health service collapse

was as imminent enough to justify the discounting of future life and wellbeing that took place. Also, if you fail to adequately prepare the health system for predictable shocks, is it rational to then pay *any* price to avoid overloading it—if not, then where is the threshold?

**Barratt.** (1) Wood and Wit (2021) treat English NHS regions separately. All regions have peak incidence before the first lockdown, but $R > 1$ for the North West at lockdown and it is certainly true that London has an earlier peak and earlier $R < 1$ point than the rest of the country. Of course this begs the question of why lockdown would be necessary in the rest of the country if the most densely populated and connected region of the UK had infections under control without lockdown? (2) We analyse aggregate dynamics because incidence is additive and the aggregate data have a higher signal-to-noise ratio. Spatial variability is indeed important: as Diggle points out, a modern surveillance system should be able to provide this. But we believe that the bigger problem remains the imbalance between the attention given to Covid risk versus the long-term risks from the response. On (3) we think that the first priority has to be well calibrated communication of risk from the pathogen. If that has been done then locally adapted risk maps are highly desirable. However, if communication amounts to mapping risk on a scale that runs from 'panic' to 'panic massively' then it is not so useful.

**Berger et al.**'s comments on the unhelpful sudden emergence of media Covid experts mirror our view of the UK, although the culture war aspects, so well caught in Julia Zeh's novel *Über Menschen*, perhaps played out differently in Germany than they did in the UK. Their demonstration of how widespread is the issue of neglected population ageing in excess death calculations is particularly interesting and important. We add this to the lessons learned!

**Britton**. We agree with the points made, particularly about the necessity of assessing the risk from negative effects of the interventions before and while they are imposed. While we agree that the scientific involvement in measurement and vaccine development in the UK was excellent, we are less convinced that the high level of involvement by many academics was entirely positive, partly because of the dynamic whereby the scientists were focussing only on Covid, and the sheer volume of their work then tended to crowd out the other societal concerns that should have been as or more important. Also, in part, we felt that the UK scientific community developed a sort of herd immunity to data or scientific arguments that challenged the dominant scientific narrative, and this was not helpful.

**Chadwick et al.** We intend no criticism of Hanlon et al. (2020) for doing useful work with the data at hand (although their results did lead to the interesting result, at some point in mid 2020, where the average loss of expected life for those who actually died of Covid in the UK was less than the Hanlon estimate, if one only used the victims' age based average expectation of further life!). The injunction to only manage on the basis of facts and data available at the time seems rather difficult in practice: is the fact of co-morbidities being binarized a fact to consider or not? Should the absence of any over 80s in the POLYMOD survey used to parameterize age structured models mean that transmission among the over 80s should be omitted from models, or what?

**Chind.** These comments seem to ignore the fact that if you spend a huge amount helping someone avoid one health risk you can not spend it helping them avoid another, and indeed that if you damage the economy enough in so doing you run the risk of causing long-term health and life loss. The NICE threshold is not used at the single patient level. It is about fair allocation of resource at the population level. If one considers the fairness and utilitarian reasons for rationing in this way it is not clear what the basis for completely abandoning these in a pandemic of this nature would be.

We assume that the comments on care workers relate to Peter Goldblatt's contribution. We thought his comments reasonable and not victim shaming.

**Dagpunar**'s suggestion is very interesting but we would caution against giving too much weight to the results. Given the many model simplifications, in the absence of proven good predictive calibration we think it is unwise to give too much credence to extrapolations from these models whether described as predictions, projections or counterfactuals. Leaving aside the important but un-modelled effects of seasonality, the fundamental difficulty is the modelling of complex, variable, ever changing social animals, as if they were molecules of gas. Take Imperial College report 9—the political difficulty of lifting a lockdown once it was imposed (and people were still dying from Covid) was something completely beyond the scope of the models, yet arguably the critical practical difficulty with imposing lockdown.

**De Nicola.** We agree completely with these comments and find it reassuring that statisticians working completely independently on this issue arrived at essentially the same conclusions on the importance of accounting for ageing. The fact that their results are so similar based on independently developed and implemented methods differing in detail, further strengthens the importance of the ageing effect. This is independently replicated science.

**Di Loro et al.** offer a cogent analysis of what went wrong with the presentation of risk and science. We were particularly interested in learning about their efforts to counter the misleading narratives disseminated to the general public during the pandemic. Although statistical literacy is important, if this discussion meeting shows something, it is clear that statistically literate people might come to different conclusions, even with the advantage of hindsight. On the other hand, discussion meetings like these hopefully mean that as a statistical community we can start to work toward a deliberative equilibrium.

**Diggle**. As we noted in our response to Barratt, we basically agree that having a public health surveillance system that can deliver predictive inferences on key health indicators at fine spatial resolution in near-real-time is a good idea.

**Dodd et al.** We note that the credible interval for the Dodd Figure 1 is compatible with our results, although we wonder if some artefact of inappropriate assumption of independence has crept in to the Ellison et al. (2021) method to generate quite such wide intervals. Note also that some of the ageing effect that we account for may not be captured by the Ellison method because of the wide age bands used. We do not think that the suggested cancellation in the ONS figures is what is happening. As we show above, allowing for recent life expectancy trends adds around 28,000 excess deaths to our figures (the use of 60 year trends in this context seems unjustified).

**Dyson**. Section 1. We agree that robust estimation of the impact of the control measures on Covid is very difficult, but this does not seem to have prevented a widespread very firmly held view that the measures were both effective and proportionate. If the Ferguson et al. projections were good enough in March 2020 to be used as justification for lockdown it is unclear why they would not be good enough for an approximate life year calculation, even with economic forecasts that substantially underestimated the actual costs. Section 2.2: We don't see how the global economic uncertainty can be separated from what individual countries did, especially those that, as Streicher et al. have pointed out, have an outsized reputation and influence in epidemiology. Did the UK government's approach to Omicron have no international impact, for example? Of course the point that different restrictions can have radically different economic impacts is exactly the point.

**Engebretsen et al.** We agree that models fitted to data and used for short term prediction are unlikely to be badly compromised by neglect of individual heterogeneity. But the same can not be said for models from highly influential groups used to predict total death tolls under different policy scenarios and feeding directly into policy. We think that heterogeneity of hospitalization risk is a different issue. While it is true that modelling groups were working very hard, we are less convinced that a seriously flawed model is redeemed by being timely. In this context, it is perhaps worth mentioning that in the interval between the acceptance and publication of Flaxman et al. (2020), Nature rejected Wood (2021), without any question as to its technical correctness.

**Fang.** It seems difficult to argue that the reductions in personal contact prior to lockdown in the UK and Sweden were caused by lockdown, given that Sweden did not lockdown, even if one is unconcerned by the effect preceding the cause. Unfortunately our Norwegian is not good enough to comment meaningfully on the Statistics Norway report (Paulsen Blytt et al., 2022, only the one page summary is in English). However, the point remains that Sweden is an example of how the pandemic could be handled without full lockdowns and the full economic and social damage that these entailed. A comparison of the change in Swedish and UK government debt levels after Covid at least suggests that the ongoing costs and opportunity costs would have been lower under a less stringent approach.

On cases and waste water, we refer the reader to Figure 2 of Fang et al. (2022).

**Fanshawe**. Certainly MD in Private Eye was wise before the event and we don't think he was alone. The point about risk is that it includes uncertainty. If you pile up poor quality studies you tend to exaggerate in one direction. Honest risk communication would not be stressing rather speculative dangers, but would, we think have been closer to the MD column from before lockdown. Point out that there is a risk of post viral complications, but also that virtually everyone

was going to get Covid eventually and it was highly uncertain that any vaccine would protect against post viral complications.

Fanshawe is right that we got it wrong about blinding in the Matta study (an example of multiple rounds of revision not always leading to improvement). We did not cover Stephenson et al. (2023) as we were not aware of it when first writing this section in early 2023. While another example of a properly conducted prospective study it is affected by the same drop out issues we highlight (as the authors acknowledge), and it is perhaps worth quoting the paper's conclusions *in full*.

> In CYP [Children and Young People], the prevalence of specific symptoms reported at time of PCR-testing declined with time. Similar patterns were observed among test-positives and test-negatives and new symptoms were reported six months post test for both groups suggesting that symptoms are unlikely to exclusively be a specific consequence of SARS-COV-2 infection. Many CYP experienced unwanted symptoms that warrant investigation and potential intervention.

We note that Fanshawe does not comment on the ONS long-term impact figures—are these too an example of us downplaying the risk and if so how?

**Fisch et al.** On point 1 see the discussion from Di Loro et al. Point 2 is covered in the main response. On point 3 we are not clear how the ONS Reff is obtained before the ONS survey started, and we could not find the relevant information on the link provided. On point 4 we note that during the apparently harsh Swiss measures the ski resorts and borders were open.

**Fisher** We completely agree with this, and especially the need for measurement to happen as soon as it becomes apparent that there might be a problem. We tried to make the measurement point early on in 2020, but were utterly naive about how to go about it, or how much noise would have to be made to have a hope of being heard.

**Goldblatt**'s, clearly heartfelt, contribution mirrors the objections of the referee whose views did most, over 4 refereeing rounds and a year and a half, to ensure that our paper would emerge too late to contribute to the national debate surrounding the Covid enquiry's consideration of the science. Taking the points in turn. Accounting for population ageing and the baby boomers is not somehow separate from the life table approach: the latter is what needs to be done to achieve the former. Fundamental to point 1 is a belief that austerity was unrelated to the financial crisis —we do not believe this is credible, but readers can decide for themselves. On point 2, there is obviously little point in computing excess deaths to assess the effects of the epidemic if you have *defined* Covid deaths as excess a priori. In addition since dying of any cause means that you die earlier than you would otherwise have done then Goldblatt's definition gets close to defining all deaths as excess. Point 3 seeks to disagree with the point we make about the effects of harvesting within a 3 year period and to support this then makes exactly the same argument for a 100 year period: it is difficult to respond to logic of this sort. On point 4 we obviously did not argue that lockdowns were unnecessary because cases and deaths are lagged data etc. With verbal models of the sort then presented you can prove anything. On point 5, to argue that self reporting exaggerates rates is not to argue that the underlying rates are zero: we do not negate. Point 6 presumably does not require rebuttal for anyone who has actually read section 3 of our paper.

**Gomes et al.**'s demonstration of the potential to leverage regional differences in initial conditions to address the identifiability issues that arise when modelling heterogeneity is particularly appealing. Combined with better measurement of individual heterogeneity in contact rates, we think that this work points the way to adequately dealing with the heterogeneity issue in future (exactly the sort of methods development that Keeling calls for).

**Gupta**'s point, that when you combine the seasonality, heterogeneity and behavioural changes and acknowledge the underestimation of exposure inherent in serology data (a point also made by Fanshawe), the turn around in infection rates before lockdown is unsurprising and may have as much to do with biology as behaviour, is important. The desire for control is human, but it repeatedly pushes us to ascribe to our own actions what nature and chance have given us little real control over.

**Hall** In response to the points we could follow: 1. Does this mean that risk distortion is justified? 3. Seems to conflate case data and random samples. See what REACT2 actually did. 4. See the time dilation approach in Wood (2021).

**Held et al.** gives a nice overview of the rather rigorous statistical approach that partly underpinned what we view as the less damaging Swiss measures. While Switzerland is of course also a rich country with a well resourced health system, we also wonder if the generally high standard of technical literacy within government may also have played a part in achieving somewhat better risk management. Purely anecdotally, in August 2021 EW spoke at the Swiss Statistics Days conference in the same session as Alain Berset, Swiss President of the Council of Ministers. With Covid still ongoing, Berset's team were very keen to get the preprint of Wood and Wit (2021), in some contrast to the work's UK reception. In contrast, during Covid the UK science and technology committee of the house of commons had 3 members with a science background and Boris Johnson's then chief advisor, Dominic Cummings, has a degree in ancient history.

**Hill** disagrees somewhat on the utility of case data. Of course this depends on what you are doing, but we believe there was far too much emphasis given to these data of convenience once actual measurements became available.

**House**. That Medley commented on Gomes on Twitter seems weak evidence of the heterogeneity effect being taken seriously. House's comments elide heterogeneity of susceptibility/transmissibility with household or network structure. This is odd given that House and Keeling (2011) take pains to avoid such confounding, by setting the degree of each node in their network models to be constant. The nice paper by Pellis et al. (2020) focuses on populations structured into children and adults and household structure, and not person-to-person variability (see their Methods section). The Dattner et al. (2021) data is on transmission within households during lockdown: its relevance to the question of heterogeneity in normal contact rates is unclear. So the cited references do not appear to suggest that the heterogeneity we discuss does not have a unidirectional effect, especially so in the context of the simple SEIR type models that were actually calibrated against data statistically.

**Ioannidis.** In all honesty we can not disagree with the points made. A public health emergency is not the moment at which to abandon the evidence based approach, but much of the response, from the untested use of lockdowns to the absurd restrictions on outdoor activity, did just that.

**Irons** is entirely correct that the alternative to the measures taken was never going to be business as usual: there would be a substantial economic hit in any case. But the point of drawing attention to the figures as we do is to emphasise that the consequences were so large that the 'spend whatever it takes' attitude would be bound to cause immense future hardship. In the UK that is evident now in the massive extra post pandemic debt servicing payments that inevitably can not be spent on health, energy transition, education or any number of other socially desirable areas.

**Johnson.** We don't really understand how distinguishing targetted and mass testing relates to the distinction between cases and prevalence. Even if one agrees that the right hand panel of Johnson's figure 2 would be an acceptable error in a derivative in this context, we don't see how the number of tests administered each week relates to the assumption that those tested are somehow a random sample of those tested, or the concern that the ascertainment fraction is likely not to be constant. The problem with high volume Twitter commentary is that it is rather easy to pick out where it turned out to be right, after the event. If we recall rightly, @BristOliver was less prescient when it came to hospital load from Omicron.

We agree with Johnson that lockdown almost certainly prevented additional Covid deaths during the first wave (whether the price was then paid in the second wave, as modelling suggested it would be, is another matter). But basing the cost of lockdown on the cost of furlough alone seems rather economically naive, and takes no account of the economically mediated future life loss, which, while hard to predict, appears from previous data to be a substantial risk.

**Jose**'s comments on risk communication are important, which is why we think that David Spiegelhalter's original explanation of risk in terms of background risk in a year was so well done, and why we concur with him that it is a pity that it did not form the basis for the approach to risk communication taken. Whatever one's level of numeracy, one is likely to have some level of understanding of how one feels about risk of dying this year and hence to gain some useful understanding. A school friend of SNW's daughter commented in autumn 2020 'the way it was presented I should have known loads of people who died from it and I don't know anyone'—that this feeling was anything but unique suggests the sort of risk communication failure that can only erode trust in science in the future. We agree too with the comments on models. What did the UK do well? Once it got going, the ONS survey that actually measured the state of the epidemic

was excellent, and we remain baffled by the failure of other countries to replicate this. REACT and REACT2 were also superb and the openness of much data was also exemplary. The central UK government not pursuing the pseudoscience of zero Covid and, perhaps through growing model skepticism, declining to lock down for Omicron were also huge positives. The UK's lab based biomedical scientists, like those of several other countries, obviously did a great job on vaccines.

**Keeling** complains that we chose to pick a comparison in which models did badly, and there were many cases where they did well. The difficulty here is that the available tests are rather asymmetric: it is easy to predict that if you suppress hard then cases will come down, and much harder to predict either the outcome under reduced measures, or how little suppression would have sufficed. But management by model requires that the model is capable of making all these predictions, not just the heavy suppression ones. We can all agree that a sledgehammer will crack the nut, but how small a hammer could we have used? The predictions for Sweden and Omicron are the real tests that we have of the ability to get close to this latter question. On hindsight and Omicron, SNW was interviewed for Scottish television at the time, pointing out that the Omicron models being used appeared to be highly pessimistic chimera, but getting heard at this point at the time was difficult (as was contributing to policy at all, unless one bought in to the dominant narrative).

Keeling asks for the lessons to be learned for the next pandemic, and we have listed some in the main points section. Keeling himself emphasizes the need for new method development, but we think that such developments are perhaps less important than the need for new measurements (and perhaps better communication of what existing methods can and can not do). For example, the POLYMOD survey (Mossong et al., 2008) used to parameterize age-specific contact rates in many epidemic models was based on 1-day diaries kept by around 1,000 UK residents of whom 7 were over 75 years and none over 80 years of age. That is a data deficit that should probably be a much higher priority for rectification than method development.

**Kumar** makes an important point about dark data, of which perhaps the most troubling early example was the WHOs presentation of case fatality rates without the appropriate level of emphasis on the dark data of undetected nonfatal cases. Another is the missing data on social contact rates between those in the most vulnerable age groups. It is possible that lockdowns had some short term life prolonging effects, of which the most obvious would be the interruption of the normal circulation of respiratory pathogens that are often among the causes of death for the very elderly. Against this must be stood the reduction in access to healthcare for those suffering from life-threatening conditions such as heart attack, sepsis, and stroke—the conditions that usually keep ICUs near capacity. At least in 2020, ONS data on deaths at home were consistent with rates for these conditions having increased.

**Liu et al.** We think that the convolution process that smooths incidence trajectories as they become sytmptom onset, hospitalisation and death trajectories calls for considerable caution in modelling using step functions aligned with policy announcements. At very least all conclusions made using them need to be checked against alternatives that assume more continuous dynamics. Similar caution is needed when assuming smooth trajectories of course, and Wood (2021) spent considerable effort on checking the effects of assuming a smooth trajectory. The same does not seem to have been true for the analyses assuming step changes. With the hindsight provided by the REACT2 and ONS incidence measurements it is at best unclear that step functions or similar are appropriate for incidence.

**Maruotti et al.** We agree completely with the need for serious consideration of the very real trade-offs and the crucial importance of the difference between good model fit and good predictive validation (as emphasised in the context of epidemic models in Wood & Thomas, 1999, for example).

**Molenberghs.** The Abrams et al. (2021) model cited as an example of models including person-to-person variability is simply a stochastic implementation of a model without person-to-person variability except that related to age (section 2.2 of Abrams gives the model), perhaps emphasising the poor understanding of this issue. Belgian debt has not risen as dramatically as in the UK, so the economic effects may well be less (debt servicing costs are another matter, of course). Earlier intervention is obviously better if all you care about is the immediate problem. But earlier interventions will presumably need to be in place for longer and create a larger problem the following respiratory virus season. Hence it is entirely unclear that they reduce, rather than increase the economic damage and consequent health and societal

damage. Particularly unclear is that this is rational risk management before it is clear that a vaccine is possible.

**Pellis**. Because of the profoundly damaging and untested nature of the intervention we would not have imposed lockdown 1, viewing the downside risks as too high and the intervention as having no viable exit strategy given the uncertainty over the possibility of a vaccine; this was especially so given the obvious political difficulty there would be in lifting a lockdown in a timely manner. However, we think this decision was finely balanced and other views were reasonable, given the uncertainties and weak health service capacity. However, what was profoundly wrong was the duration of the lockdown, which was irrational unless the scientific fantasy of zero Covid was being pursued.

The deconvolution analyses were first produced in April 2020, in plenty of time to have been used in the management of the last pandemic, let alone the next, but then, as now, there was a reluctance to take the implications on board.

The deconvolution method is not a GAM. One can do an approximate version of it using GAMs, but this involves assuming smoothness on the raw incidence scale, which is obviously suboptimal. The fact that a method is mechanism agnostic does not mean that it can not detect a pattern generated by a mechanism, nor diminish the utility of detecting it as a model sanity check (notice that the MRC method also detects the pattern).

We are not sure what useful purpose is served by Pellis' Figures 1 and 2. The smooth in Figure 1 is obviously the wrong model for the data presented, so the $R$ estimates resulting from it will obviously not be correct.

We completely agree that fine control of fast-growing systems with long delays is really hard. Doing it on the basis of mathematical models never designed or validated for the task is even harder. That is why we think that in the next pandemic early measurement should be the biggest priority: the testing capacity was there this time.

**Rougier**. The first lockdown is where the distribution of time to death is likely to be most reliable, given that the distribution was estimated from fatal hospitalized cases up to October 2020. The exception to this is that we used the distribution from time to infection to death for the general population and there is evidence that this period is longer in the older patients making up most of the deaths (Tan et al., 2020). This may partly explain why our peak is a little later than the REACT2 based estimates. Later on we think that the point is of more concern: by the time selective vaccination is changing the demographic profile of the victims there is obviously a worry, even without variant changes. But by then there are direct sources of information for the UK, which suggest that the approach is at least not estimating peak incidence to be earlier than occurred in the general population.

The second suggestion is very elegant, but we think that the 'all excess deaths are from Covid' assumption is difficult. Paraphrasing a medic from a large ICU, to the first author in 2020: 'it was tough, we were full with Covid patients. But the thing is those beds are not normally empty, they are full of patients with sepsis, heart attacks and stroke. I've talked to colleagues all over the country and it's the same. The question is where have those other patients gone? The assumption is they are dead'. The ONS excess deaths at home data at the time did not contradict this.

**Selby**. On $R$, Wood (2021) includes more sources of uncertainty, but we agree that outside London it is unclear if $R$ was below 1 before lockdown (or even if discussing the instantaneous value of a quantity playing out over future time makes sense, at least with regard to relative timing differences well below the generation time). Perhaps there is the sort of backwards in time causality from lockdown suggested, perhaps people had a last minute burst of activity before being locked down, or perhaps, like the authors, most people really didn't know what the prime-minister would actually announce until he did so. Behaviour that changes $R$ instantly, changes incidence instantly. We're not sure what was wrong with the qualitative prediction of the models that higher first wave suppression would lead to larger second waves.

**Senn** is right—when we went back to the data on which the Marmot report plot is based it does appear that there is an error, as he points out (although it does not change the conclusions). We should have gone back to the original data before. This is in fact the only point at which we used 'second hand' data. Other data used were either from data files provided by the stated source, or sometimes digitized from the data producers own publications.

**Singh.** On long Covid we think it prudent to at least adjust the quoted prevalences using the results from the studies looking at the proportion of self reported long Covid against seroprevalence. We agree that the discussion of economic effects is difficult, uncertain and crude, but none of those things are reasons to discount the effects in favour only of the effects presented as being more quantitatively accessible.

**Spiegelhalter.** We agree with these comments and especially the need for intervals to reflect realistic calibrated uncertainty. If defensible levels of uncertainty accompanied epidemic model predictions then there would perhaps be less tendency to down-weight the highly uncertain downside risks when balancing policy options. See also the response to Jose.

**Srakar** asks what we need in order to be better prepared in future. Infection surveys ready to go for a preprepared cohort. Much stronger training in the fundamentals of what makes the statistical approach valid, so that we do not become blinded by methods into forgetting the data.

**Streicher and Broadbent** have been modest in promoting their own work, but of the many papers cited by discussants, Streicher et al. (2025) is one of the ones that most repays reading. Particularly salient is their cogent argument that $R < 1$ is simply the wrong target in practice. Their other work highlighting, for example, the inequities of the pandemic responses outside the first world is also worth seeking out. It is also relevant to the UK, given, as they argue, the disproportionate role of UK groups and institutions in framing and setting the global response. For example, when Neil Ferguson, of Imperial College's highly influential WHO collaborating Department of Infectious Disease Epidemiology, realized that it was possible to 'get away with'[10] lockdown in the UK, it was likely to have had worldwide effects well beyond the UK.

**Tian et al.** The wide versus narrow caliber issue is indeed a serious one. In the UK context death certification guidelines were changed early in the first wave to include Covid on death certificates if it was suspected that it could be involved: as one medic put it in the context of care home deaths: 'now everyone dies of Covid, previously we didn't speculate on the specific pathogen if someone dies of pneumonia, now it's all Covid'. Carl Heneghan successfully campaigned to get the the UK government to move from an absurdly wide definition that would have counted a road death as Covid if the victim had ever had Covid, but the variability even in the definition of Covid death is still a problem. In response to point (2), we are not sure why the $t$ distribution is deemed incorrect here. Code is provided for fitting the models in the supplementary material. On (3), see the contribution from Gomes and associated papers.

**Tildesley.** It is slightly surprising that those who express most scepticism about the possibility of predicting the no-lockdown counterfactual are the most convinced that lockdown was necessary. The verbal model that the vulnerable will peak before the rest of the population is difficult to reconcile with the ONS or REACT2 results which suggest that the fatal incidence peak was after the general incidence peak. The latter phenomenon is perhaps expected. We used the infection to symptom distribution for the general population, whereas there is evidence (Tan et al., 2020) that older people actually have significantly longer incubation periods: this effect would tend to shift the inferred fatal peak to later than it should be. Additionally, at least for SEIR type dynamics, sub-populations with lower contact rates tend to peak later than those with higher rates, not earlier, for reasons related to those discussed in Section 4.1 of our paper.

**Vencalek** provides a very nice case study illustrating how a pandemic does not in anyway remove the considerable difficulties that attend observational data, and the need, if anything, for extra care in the pandemic situation.

**Whitehouse et al.** are surely right to identify problems of identifiability as a central difficulty in dealing with the substantial effects of heterogeneity on disease models, although, as Gomes shows elsewhere in the discussion, the difficulty can be overstated. But whatever the level of practical difficulty, the question is how to fill the information gap in future? A start would be to devote substantial effort to measurement of the heterogeneity in contact rates. The UK models relied heavily on the POLYMOD survey: a stupendous multi-country effort to look at contact rates, but for the UK still based on around 1,000 1-day contact diaries, with 7 participants over 75 years and none over 80 years of age (Mossong et al., 2008). Given the risk profile of Covid and the seriousness of

[10] In a December 2020 interview with Tom Whipple in the Times, Ferguson is quoted as saying, with reference to lockdown in Wuhan: *'It's a communist one-party state' we said. 'We couldn't get away with it in Europe,' we thought. And then Italy did it. And we realised we could.*

the situation, this is clearly inadequate (patching a nuclear reactor with ductape?). A much larger statistical study is needed. Even then the question of how to adapt the measured baseline rates as behaviour changes in the face of risk remains as a challenge.

**Wyse et al.** Obviously the deconvolution approach is retrospective, but the analysis from section 5 was first applied at the end of April 2020. The results about peak timing were the same as now. The paper was rejected without review by JRSSC in early May 2020 for (perhaps disingeneously) 'insufficient methodological novelty'. For the incubation period the first version used Lauer et al. (2020), which was online 10 March. MacAloon's paper was on medRxiv in April 2020. So while it is true that the analysis is retrospective, it was in practice available quite early enough to have contributed to some adjustment of the dominant narrative and to policy, had there been willingness to take its message on board: instead there was a tendency for the results to be dismissed as a 'right wing talking point'.[11] See also Figure 2 of Wood (2025) for an investigation of how soon reliable inference could feed policy. See the main response for some of the other improvements we suggest. On sampling, the ONS and REACT surveys show how to do it, but it would be a good idea to have preapproved studies ready to go at the first hint of pandemic possibility.

*Conflict of interest:* None declared.

## References


Abrams S., Wambua J., Santermans E., Willem L., Kuylen E., Coletti P., Libin P., Faes C., Petrof O., Herzog S. A., Beutels P., & Hens N. (2021). Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. *Epidemics*, *35*(163), 100449. https://doi.org/10.1016/j.epidem.2021.100449

Bai J., Shi F., Cao J., Wen H., Wang F., Mubarik S., Liu X., Yu Y., Ding J., & Yu C. (2020). The epidemiological characteristics of deaths with COVID-19 in the early stage of epidemic in Wuhan, China. *Global Health Research and Policy*, *5*(1), 1–10. https://doi.org/10.1186/s41256-019-0129-8

Birrell P., Blake J., Van Leeuwen E., Gent N., & De Angelis D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: The first wave. *Philosophical Transactions of the Royal Society B*, *376*(1829), 20200279. https://doi.org/10.1098/rstb.2020.0279

Cummings D. (2023, October). *Statement of evidence to covid inquiry*. https://covid19.public-inquiry.uk/wp-content/uploads/2023/10/31180752/INQ000273872.pdf.

Dattner I., Goldberg Y., Katriel G., Yaari R., Gal N., Miron Y., Ziv A., Sheffer R., Hamo Y., & Huppert A. (2021). The role of children in the spread of COVID-19: Using household data from Bnei Brak, Israel, to estimate the relative susceptibility and infectivity of children. *PLoS Computational Biology*, *17*(2), e1008559. https://doi.org/10.1371/journal.pcbi.1008559

Ellison J., Hilton J., Dodd E., Forster J., Smith P., & Bijak J. (2021). Modelling weekly mortality incorporating annual trends. In *International Population Conference 2021*.

Fang Z., Roberts A. M., Mayer C.-D., Frantsuzova A., Potts J. M., Cameron G. J., Singleton P. T., & Currie I. (2022). Wastewater monitoring of COVID-19: A perspective from Scotland. *Journal of Water and Health*, *20*(12), 1688–1700. https://doi.org/10.2166/wh.2022.082

Flaxman S., Mishra S., Gandy A., Unwin H. J. T., Mellan T. A., Coupland H., Whittaker C., Zhu H., Berah T., Eaton J. W., Monod M., Imperial College COVID-19 Response Team, Ghani A. C., Donnelly C. A., Riley S., Vollmer M. A. C., Ferguson N. M., Okell L. C., & Bhatt S. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, *584*(7820), 257–261. https://doi.org/10.1038/s41586-020-2405-7

Hanlon P., Chadwick F., Shah A., Wood R., Minton J., McCartney G., Fischbacher C., Mair F. S., Husmeier D., & Matthiopoulos J. (2020). *COVID-19—Exploring the implications of long-term condition type and extent of multimorbidity on years of life lost: A modelling study*. https://wellcomeopenresearch.org/articles/5-75.

House T., & Keeling M. J. (2011). Epidemic prediction and control in clustered populations. *Journal of Theoretical Biology*, *272*(1), 1–7. https://doi.org/10.1016/j.jtbi.2010.12.009

Kandiah J., Van Leeuwen E., Birrell P. J., & De Angelis D. (2025). Contact data and SARS-COV-2: Retrospective analysis of the estimated impact of the first UK lockdown. *Journal of Theoretical Biology*, *610*(12), 112158. https://doi.org/10.1016/j.jtbi.2025.112158

Lauer S. A., Grantz K. H., Bi Q., Jones F. K., Zheng Q., Meredith H. R., Azman A. S., Reich N. G., & Lessler J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed


[11] The Google scholar record for the preprint changed from pointing to the arXiv to the 'Daily Stormer' a couple of weeks after posting (slightly surprising given that Google had apparently stopped indexing the 'Daily Stormer' some years previously, or perhaps not, given what is written in Matt Hancock's diaries).

cases: Estimation and application. *Annals of Internal Medicine*, *172*(9), 577–582. https://doi.org/10.7326/M20-0504

Mossong J., Hens N., Jit M., Beutels P., Auranen K., Mikolajczyk R., Massari M., Salmaso S., Tomba G. S., Wallinga J., Heijne J., Sadkowska-Todys M., Rosinska M., & Edmunds W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, *5*(3), e74. https://doi.org/10.1371/journal.pmed.0050074

Msemburi W., Karlinsky A., Knutson V., Aleshin-Guendel S., Chatterji S., & Wakefield J. (2023). The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, *613*(7942), 130–137. https://doi.org/10.1038/s41586-022-05522-2

Paulsen Blytt J., Bougroug A., & Sletten P. (2022). *Økonomisk utvikling gjennom Covid-19: En oppdatert sammenligning av Norge, Sverige og Danmark* (Technical report). Statistics Norway.

Pellis L., Cauchemez S., Ferguson N. M., & Fraser C. (2020). Systematic selection between age and household structure for models aimed at emerging epidemic predictions. *Nature Communications*, *11*(1), 906. https://doi.org/10.1038/s41467-019-14229-4

Stephenson T., Pinto Pereira S. M., Nugawela M. D., McOwat K., Simmons R., Chalder T., Ford T., Heyman I., Swann O. V., Fox-Smith L., Rojas N. K., Dalrymple E., Ladhani S. N., & Shafran R., & CLoCk Consortium (2023). Long COVID—Six months of prospective follow-up of changes in symptom profiles of non-hospitalised children and young people after SARS-CoV-2 testing: A national matched cohort study (The CLoCk) study. *PloS One*, *18*(3), e0277704. https://doi.org/10.1371/journal.pone.0277704

Streicher P., Broadbent A., & Hellewell J. (2025). The need for methodological pluralism in epidemiological modelling. *Global Epidemiology*, *9*(6), 100177. https://doi.org/10.1016/j.gloepi.2024.100177

Tan W., Wong L., Leo Y., & Toh M. (2020). Does incubation period of COVID-19 vary with age? A study of epidemiologically linked cases in Singapore. *Epidemiology & Infection*, *148*, e197. https://doi.org/10.1017/S0950268820001995

van Valburg M. K., Termorshuizen F., Brinkman S., Abdo W. F., van den Bergh W. M., Horn J., van Mook W. N., Siegerink B., Slooter A. J., Wermer M. J., Geerts B. F., & Arbous M. S. (2020). Long-term mortality among ICU patients with stroke compared with other critically ill patients. *Critical Care Medicine*, *48*(10), e876–e883. https://doi.org/10.1097/CCM.0000000000004492

Wood S. N. (2021). Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? *Biometrics*, *78*(3), 1127–1140. https://doi.org/10.1111/biom.v78.3

Wood S. N. (2025). Simple simulation based reconstruction of incidence rates from death data. *Biometrics*, *81*(3), ujaf088. https://doi.org/10.1093/biomtc/ujaf088

Wood S. N., & Thomas M. B. (1999). Super–sensitivity to structure in biological models. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *266*(1419), 565–570. https://doi.org/10.1098/rspb.1999.0673

Wood S. N., & Wit E. C. (2021). Was $R < 1$ before the English lockdowns? On modelling mechanistic detail, causality and inference about Covid-19. *PLoS One*, *16*(9), e0257455. https://doi.org/10.1371/journal.pone.0257455

https://doi.org/10.1093/jrsssa/qnaf124
Advance access publication 23 July 2025