To be read before The Royal Statistical Society at the Discussion Meeting on 'Some statistical aspects of the Covid-19 response' to be held in London on Thursday, 10 April 2025, the President, Professor Sir John Aston, in the Chair, , 1–32
doi:
Journal of the Royal Statistical 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, UK, [2]Institute of Computing, Univerità della Svizzera italiana, Lugano, Switzerland and [3]College of Medicine and Veterinary Medicine, University of Edinburgh, UK

*Corresponding author. simon.wood@ed.ac.uk

FOR PUBLISHER ONLY Received on Date Month Year; revised on Date Month Year; accepted on Date Month Year

Abstract

This paper discusses some statistical aspects of the U.K. 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 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 for full stay at home lockdowns having been necessary to reverse waves of infection is examined, with new analyses provided for a number of European countries.

Key words: lockdown, Covid models, risk management, risk presentation, long Covid, excess deaths, transmission heterogeneity, calibration

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 hours 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 economic activity is obviously inflationary (e.g.

[1] time of writing May 2023.

Wolf, 2023; Hall, 2009, Ch.1) even without the supply chain problems (e.g. Hutton and 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. Shine, 2022; Richardson et al., 2023) 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 historic 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 non-pharmaceutical 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 versus 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 non-negligible 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 a supplementary 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 two years of the pandemic this is likely larger than the Covid risk to the woman pictured.

[2] to quote former BoE chair Mervyn King (BBC, 23/10/2022): '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.

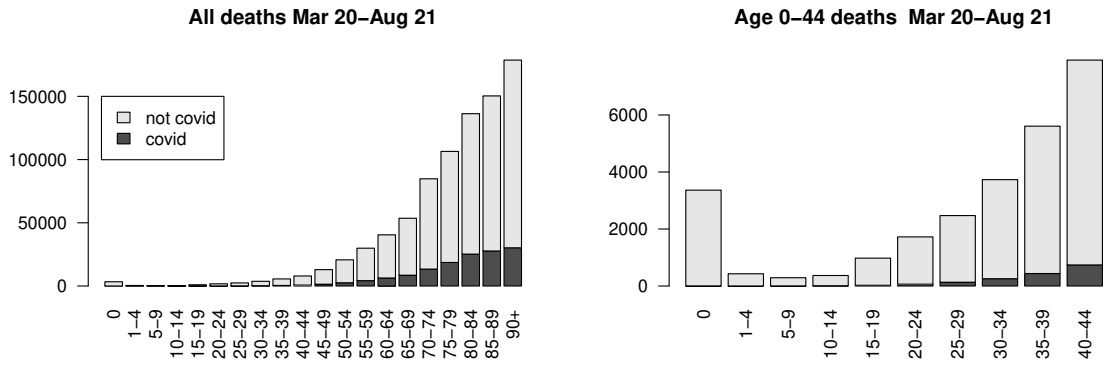

Fig. 1: All UK deaths from March 2020 to August 2021 by age band. Covid deaths are shown in black. Data are from the Office for National Statistics. The right panel simply enlarges the plot for the under 45s, where the risks are otherwise too low to be visible.

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. Hopkins et al., 1999; Herridge, 2011), hospitalised patients clearly suffered long term effects (Ayoubkhani et al., 2021, for example[5]), and a novel virus was always likely to result in an increase in people suffering longer term post-viral complications (e.g. Appelman et al., 2024).

[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.

However, beyond anosmia (loss or change in 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.

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 non-representative samples, with no control group, so that a meta-analysis of 174 studies produced up to Janaury 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 et al., 2021) found prevalence of long Covid symptoms at 5% in confirmed Covid cases versus 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 381/1782 Covid cases with at least one ongoing symptom at 90-150 days post infection, versus 361/4130 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 drop out was high. The initial Covid cohort size was 4,231, so that there is considerable scope for differential drop out 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 (as done by Matta et al., 2022, for example). 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 non-hospitalized patients, or even to rule out these being zero.

In fact, for non-hospitalized cases, carefully designed studies often found low, or even no, difference 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 non-hospitalized cases and test-negative controls. In a large population-based French cohort for ages 18-69 (Matta et al., 2022), self-reported Covid-19

[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 drop out in the fully recovered is also not implausible.

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 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 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 suggested also 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 pre-pandemic 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 program (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 program were often framed in terms of 'saving life versus 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 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, Fig 2.3). Taking the upper 10% as representing 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

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

[8] see, e.g. p. viii 2010 Conservative party manifesto, p14 2010 Liberal Democrat Manifesto

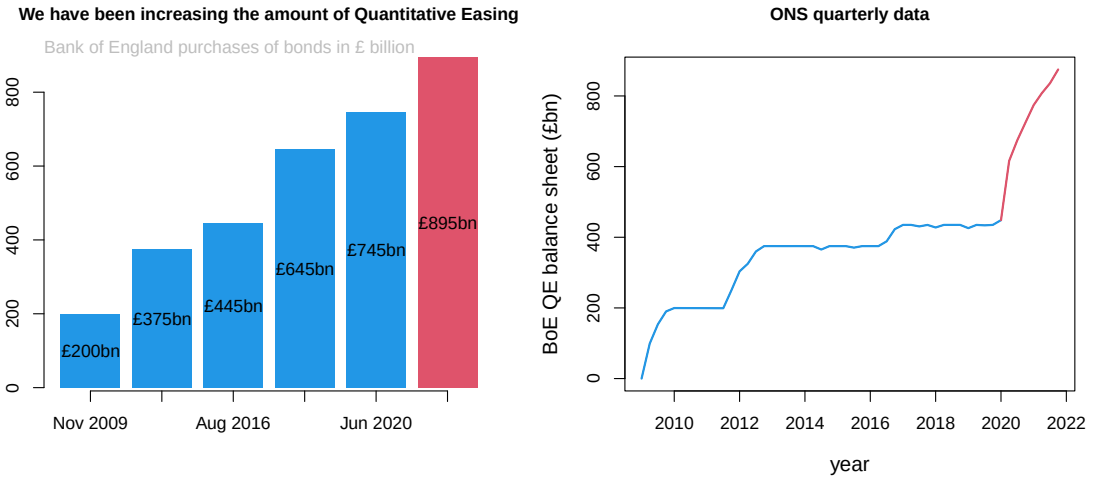

Fig. 2: Left: The Bank of England quantitative easing (QE) program as presented on the Bank of England website in 2021 (redrawn, but identical in all relevant respects). Right: the same QE program using a continuous linear time scale (ONS data), with the Covid part shown in red.

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 (Ruminska-Zimny, 1997; Ciment, 1999; Case and Deaton, 2017). 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 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.

[9] They found a 1.2% reduction in all cause mortality for each 100USD per capita increase in welfare spending. The increase of around £1000 per head in annual government debt servicing costs since 2020 is obviously unavailable for increased welfare.

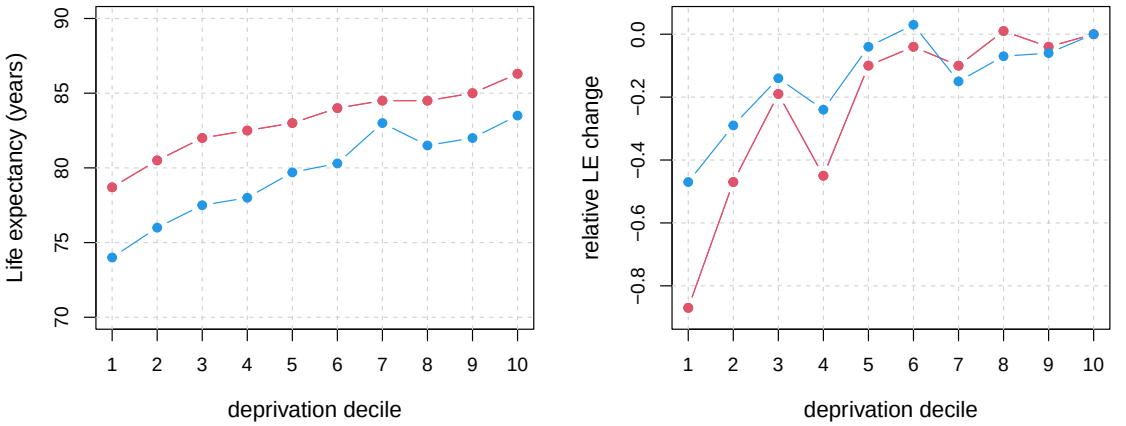

Fig. 3: Left: Life expectancy by (population weighted) area deprivation decile for women (red) and men (blue). Right the change in years of life expectancy relative to the least deprived between 2010-12 and 2016-18 for women (red) and men (blue). Data from figures 2.3 and 2.5 of Marmot et al. (2020), which were based on ONS and PHE data.

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 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 historic 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, post shock. Although the detailed mechanisms are of course complex, the ratio of standardized mortality rates in social classes V to I had actually increased in 1949-53 relative to the pre-war 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 et al. (1998) report were, may at least partly reflect the difference between what was politically possible in the recession of the early 1980s versus 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 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 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 one 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 one 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[11]

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 7260 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 was 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 versus 6 days is quite substantial. Presenting a figure in a way likely to lead to a 60 fold over-estimation 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

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

[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.

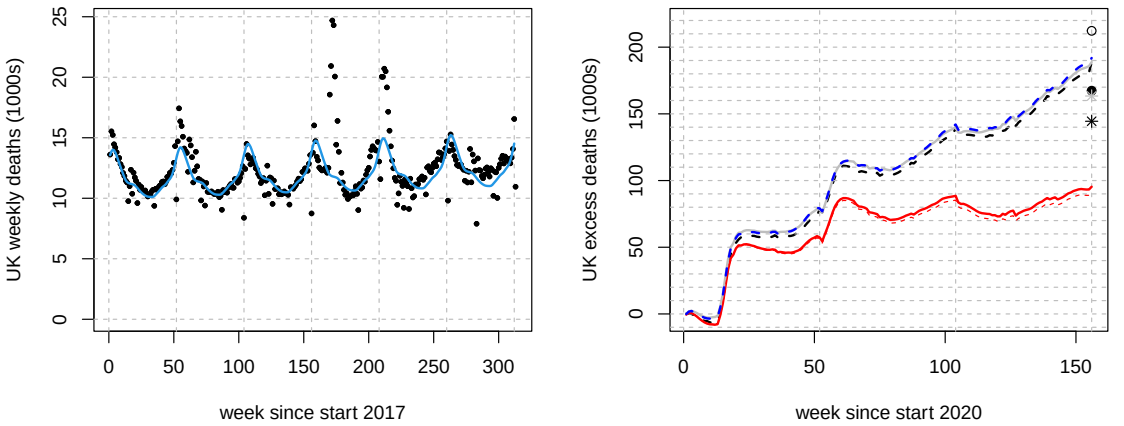

Fig. 4: Left: Weekly UK deaths (black dots), and weekly deaths predicted using 2017-19 life tables (blue). The predictions start from the age structure of the population at the start of 2017 for the first 3 years, and the age structure at the start of 2020, thereafter. The cumulative observed and predicted deaths match to within 0.0025% (50 deaths) over 2017-19. 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 2017-19 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. 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 pro-rata 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.

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 post-war 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 2017-19 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 GAM to the reference period data with a smooth for week, and a cyclic smooth for week of the year. From this, a multiplier relating 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 five 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 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 number of deaths per year that ageing has caused. The total is about 30 thousand expected extra deaths per year and the right hand plot of cumulative changes in death with age illustrates how the change is accumulated across the age groups. Note that migration makes 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 2017-19 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 emphasise 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. 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 co-morbidities are not independent events, Covid mortality events do not simply add to the mortality caused by the other co-morbidities. 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 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.

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 is 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 E, but

[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.

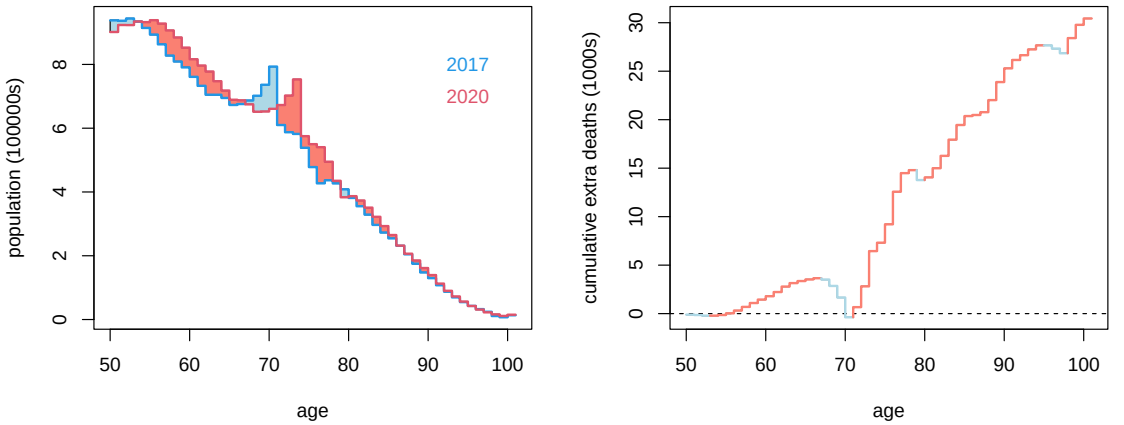

Fig. 5: The effects of population ageing on deaths. Left: blue is the UK population by one year age band in 2017 and red is the equivalent in 2020. The difference between the two (Δ_i for age class i) is shaded blue when the 2020 population is less than the 2017 population and salmon when the 2020 population is larger. Right: the population difference, Δ_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 three 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.

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 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 co-morbidities as simple categories (often binary), with limited possibility for incorporating co-morbidity severity, leading to likely inflation of the estimated life loss per victim. For example, suppose that each victim with congestive heart disease as a co-morbidity is assigned the average life expectancy of someone with congestive heart disease. Then we will over-estimate 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

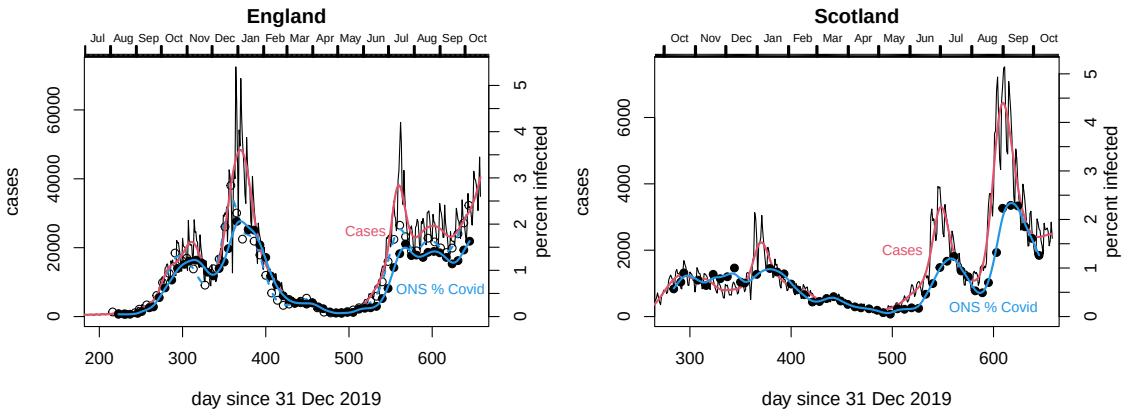

Fig. 6: Comparison of official daily Covid cases data (black lines, red smooth) and ONS prevalence measurements (black dots, blue smooth) for England on the left and Scotland on the right. In both cases the data are scaled to match over the 50 days from the midpoint day 430. If case data reliably measured 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.

to by track 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 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 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 track and trace, which apparently increased it. It seems improbable that

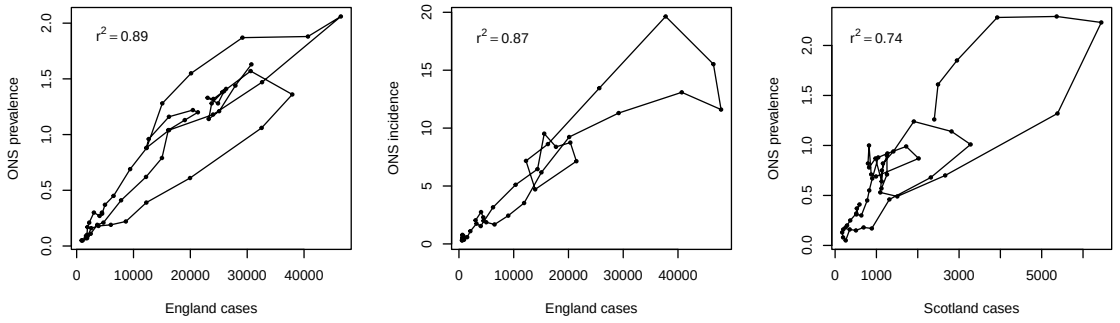

Fig. 7: Left: ONS prevalence plotted against cases (weekly cycle smoothed out) for England, with points joined in time order. Middle: similar, with ONS incidence on the vertical axis. Right: as left but for Scotland. The shifting nature of the relationship between cases and prevalence or incidence is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

track 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 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 4 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).

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

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

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 the 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 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 and 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 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 α be a parameter determining susceptibility to infection, which varies over the susceptible population, and let $s(\alpha, t)$ denote the susceptible population per unit α interval with parameter α at time t. Without loss of generality we can scale the problem so that the initial population is 1, in which case $s(\alpha, 0)$

[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`

is the initial p.d.f. of α. 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 $I(t) = \int i(\alpha,t)d\alpha$. On integrating the first 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 α. *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 α

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

where M is the moment generating function of the initial distribution of α (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 α.

Variability in contact rates can be modelled in a similar way. It is assumed that transmission depends on the product of α for the susceptible and α' 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 α 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 $\bar{\alpha}'_t = \int \alpha^2 s(\alpha,t)d\alpha / \int \alpha s(\alpha,t)d\alpha$. 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.

If α has a gamma(k,ν) distribution in the initial susceptible population, with p.d.f. $\nu^k\alpha^{k-1}e^{-\alpha\nu}/\Gamma(k)$, 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 α 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 λ 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 λ 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.

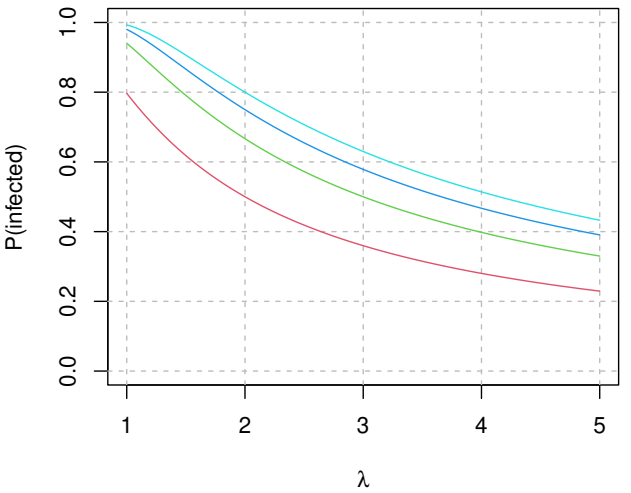

Fig. 8: Final proportion infected against λ for $R_0 = 2, 3, 4, 5$ (ascending, red, green, blue, turquoise), for SEIR epidemic models with susceptibility or mixing rates varying across individuals. The age band dependent variability used by SAGE models in the UK corresponds to $\lambda \approx 1.2$. Realistic estimates seem to be in the range 2.5-5.

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 are in the key-worker population, so that over time the 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

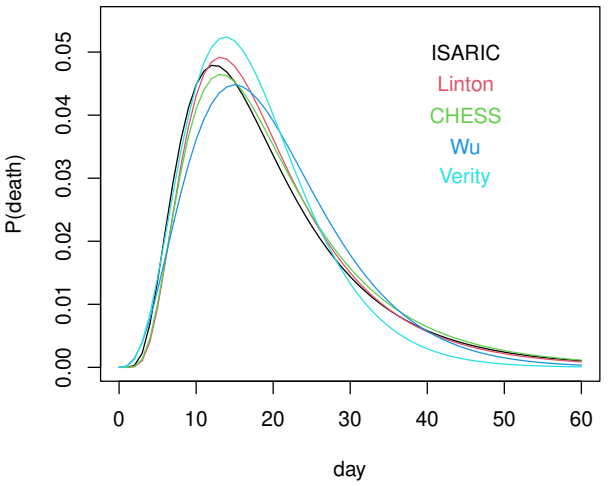

Fig. 9: Comparison of onset to death duration distributions from various sources. The ISARIC distribution is used here, because it is based on by far the largest sample size. Note that the distributions are not significantly different given the sample sizes involved.

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 can not 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

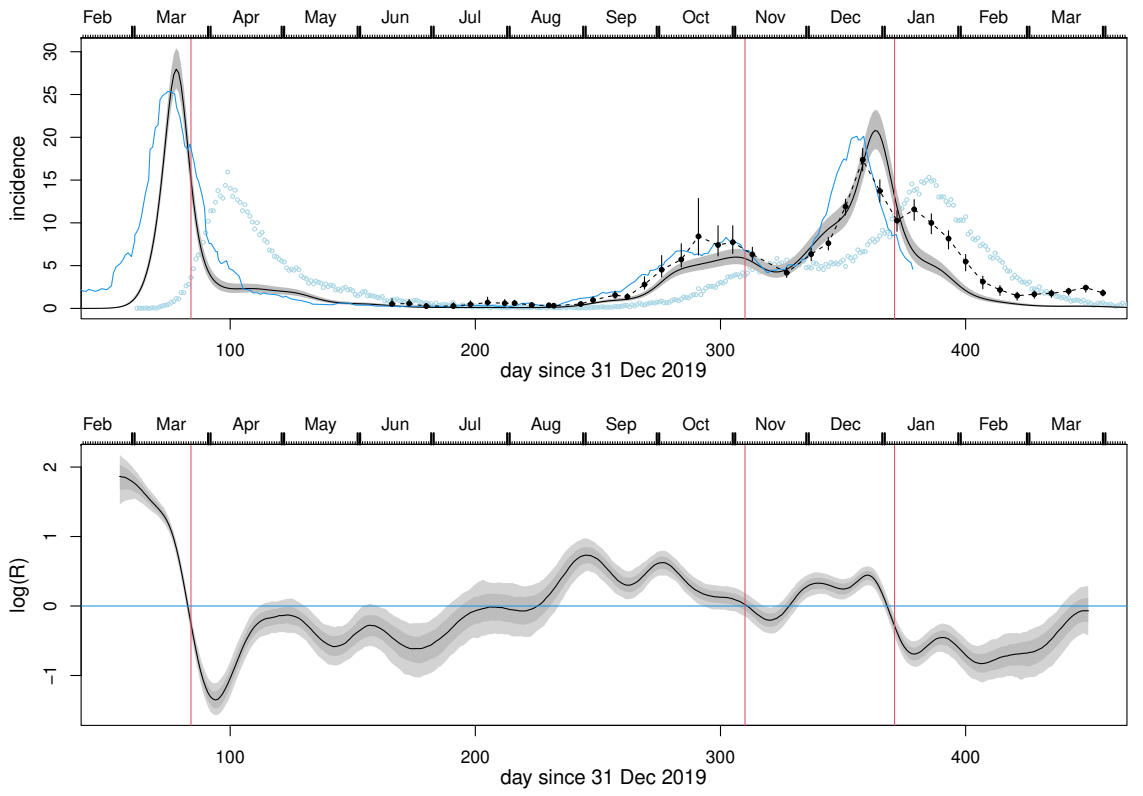

Fig. 10: Top: The grey bands are 95% credible intervals for fatal incidence (new infections per day) per million, reconstructed from the NHS England daily hospital death data shown as light blue circles. The dark blue curve shows (scaled) reconstructed incidence from the REACT-2 study's random sample 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.

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 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 McAloon et al. (2020) results, the corresponding infection to death duration model is $\log(d) \sim N(3.151, 0.469^2)$, and given the large underlying sample size, 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 (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 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 pre-lockdown behaviour, there are not enough of them to alone account for the overall pre-lockdown 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. Sub-populations 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

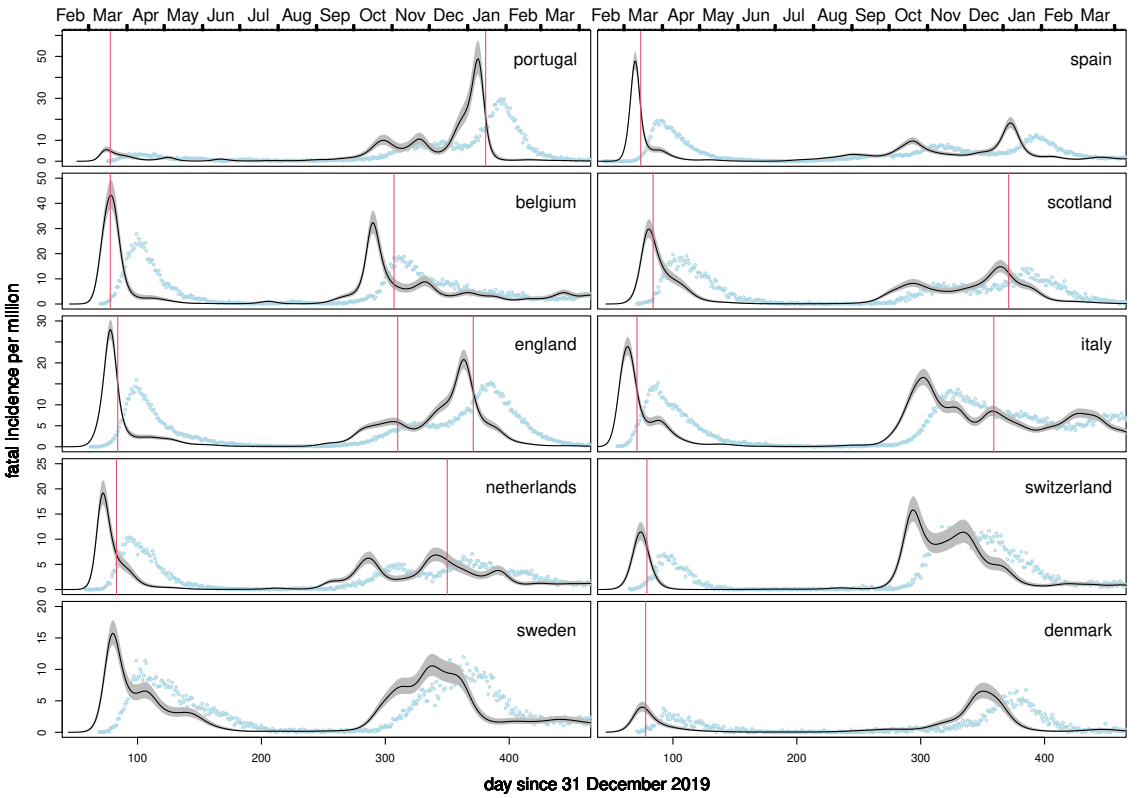

Fig. 11: Reconstructed fatal incidence for the 10 countries for which reliable daily Covid deaths data by exact day of death were available. Grey confidence bands show reconstructed fatal incidence per million population, with blue circles being the daily deaths per million from which they are obtained.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

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 the light of early model 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 supplementary material for lockdown dates and data sources.

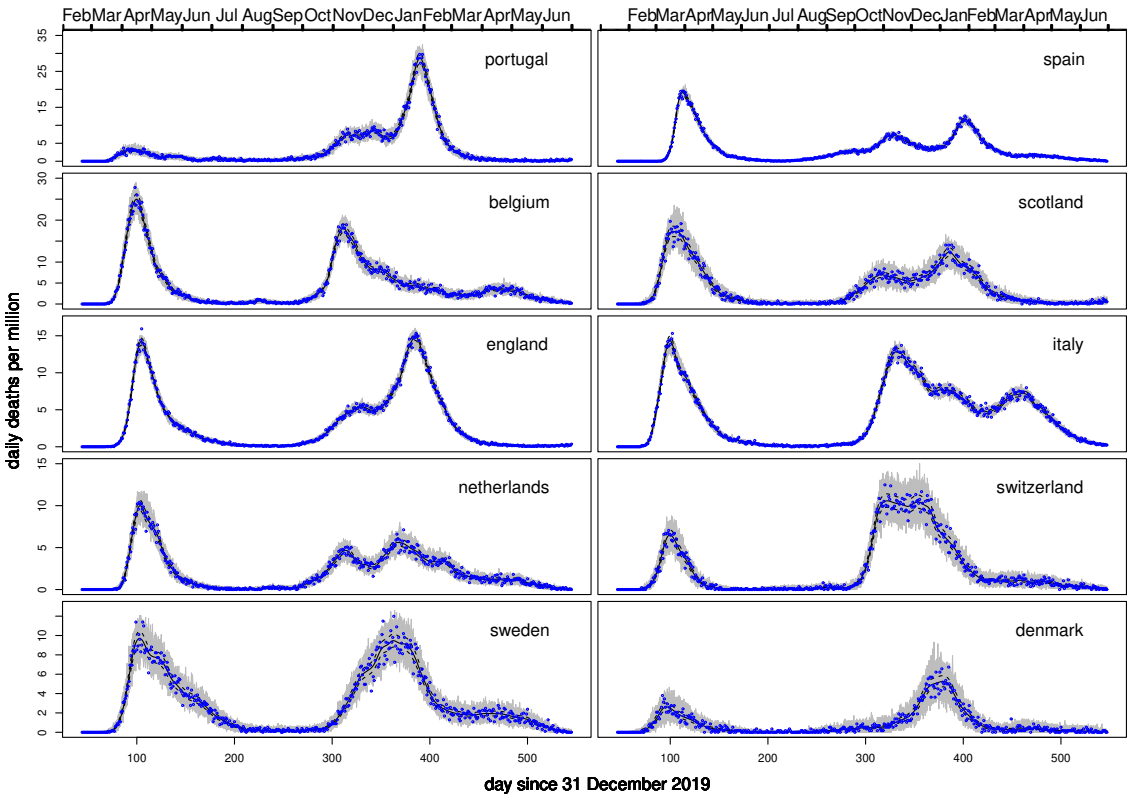

Fig. 12: Checking plots. Each inferred fatal infection was randomly assigned a duration from the fatal disease duration distribution, to produce a daily death curve implied by the fitted model. This simulation process was repeated 100 times to give the grey curves, which can be compared to the observed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

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.

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

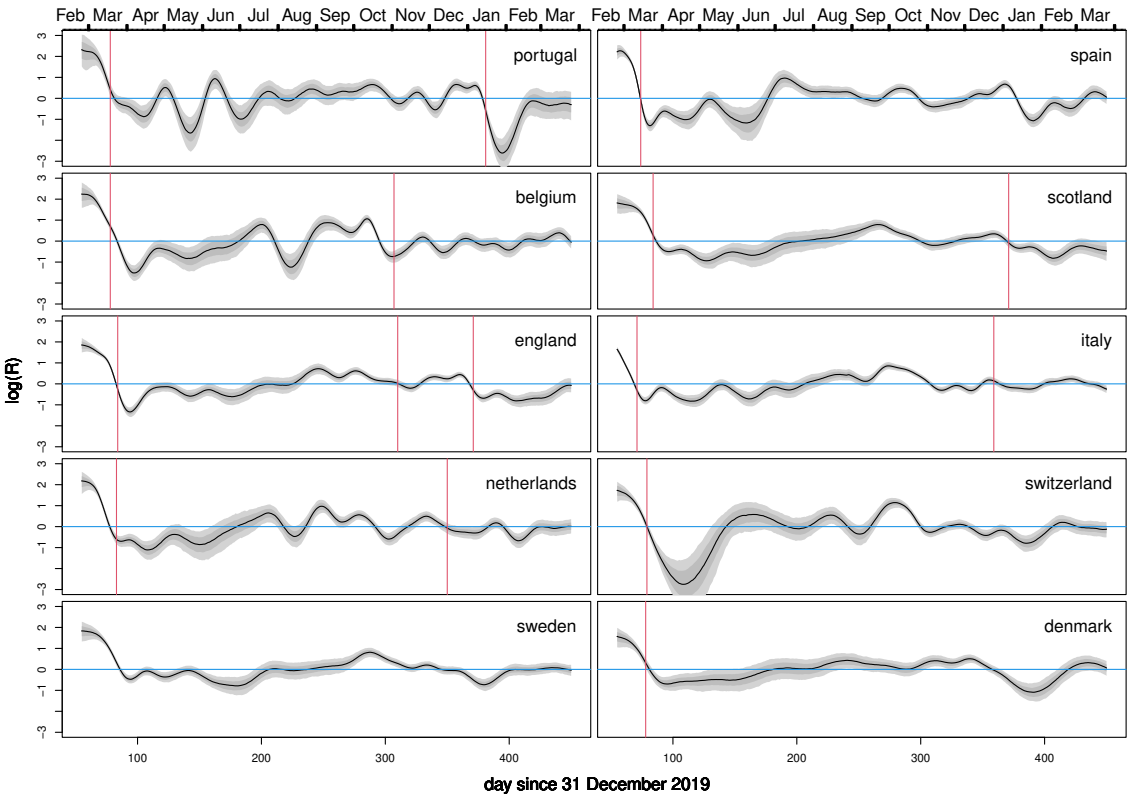

Fig. 13: The $\log(R)$ trajectory required for a simple SEIR model to produce the reconstructed incidence curves, assuming a mean time to infectivity of 3 days and a mean infectious duration of 5 days. Again vertical red lines mark the lockdown dates. The reconstructions are conditional on an SEIR model 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.

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,

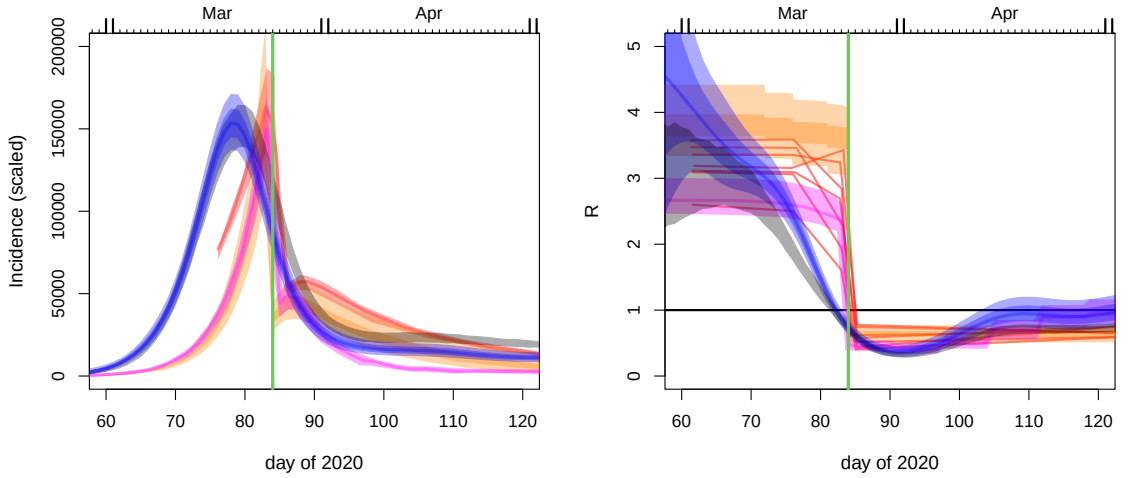

Fig. 14: Left: incidence reconstructions (vertically scaled for plotting). Right: R reconstructions. The green vertical line is the first day of lockdown. Red: IC Knock et al. (2020) (R separate for each region); Orange: IC Flaxman et al. (2020); Pink: MRC for London Birrell et al. (2021); Grey: replication 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.

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 Sweden did not lock down, the final intervention in Sweden was modelled as if it was lockdown. The approach has been widely criticised (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 can not 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 non-linear model, so Wood (2021, see also Appendix B) re-implemented 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 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 7 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 post lockdown dip and recovery in R expected if lockdown reduces

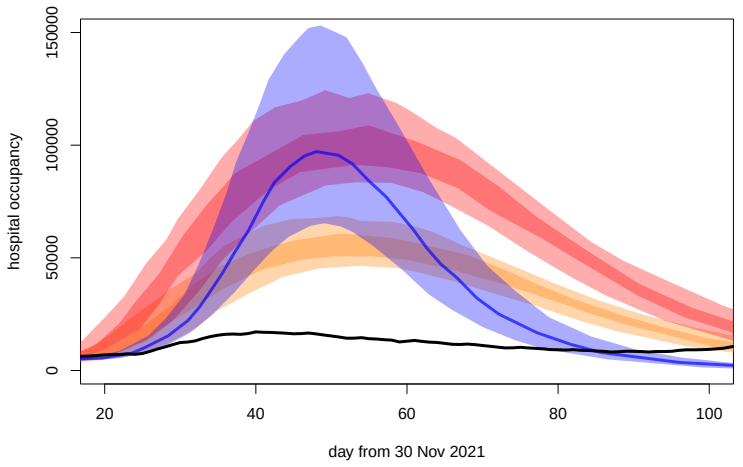

Fig. 15: LSHTM projections for NHS England Covid hospital bed occupancy as a result of Omicron, without lockdown in orange and red. These are the scenarios that are most optimistic about booster efficacy, with the orange representing low vaccine escape assumptions and the red high vaccine escape. In 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.

contact rates. Wood and 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 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 Scientific Advisory Group for Emergencies (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 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 non-linear 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 reality promoted by the online environment provides fertile ground for the post-modern 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 can not 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 6000 per day by March 20th, up from 1500 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 hours 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 25th 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 19th 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. SNW is also grateful for help from and discussions with a number of medics who cannot be named, and Dr Alistair Montgomery who can be.

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 2017-19 data, where f_1 is a cyclic smooth function and f_2 is a centred slowly varying smooth function, while σ and ν are parameter 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 one 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 one year age group $a = 0, \ldots, 100$. The average proportion of the age group then dying in one week is $q_a = 1 - \exp(-m_a/52)$. Hence the proportion of one 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 one 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 supplementary material.

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 β_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 $\lambda \int f''(t)^2 dt = \lambda \boldsymbol{\beta}^\mathsf{T} \mathbf{S} \boldsymbol{\beta}$ ($\mathbf{S}$ known), which can be used to penalize complexity of f during inference, tuneably via the smoothing parameter λ. 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

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmax}}\ l(\boldsymbol{\beta}) - \frac{\lambda}{2} \boldsymbol{\beta}^\mathsf{T} \mathbf{S} \boldsymbol{\beta}, \tag{1}$$

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

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

where $\mathbf{V}_\beta = (\partial^2 l / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\mathsf{T} + \lambda \mathbf{S})^{-1}$. Writing $\pi_G(\boldsymbol{\beta}|\mathbf{y})$ for this Gaussian approximation to the posterior and $\pi(\boldsymbol{\beta})$ for the smoothing prior, the marginal likelihood is approximately $\exp\{l(\hat{\boldsymbol{\beta}})\} \pi(\hat{\boldsymbol{\beta}}) / \pi_G(\hat{\boldsymbol{\beta}}|\mathbf{y})$ (Laplace

approximation) which can be maximized to estimate λ. This approach is equally applicable when there are several smoothing parameters and a penalty of the form $\boldsymbol{\beta}^\mathsf{T}\mathbf{S}_\lambda\boldsymbol{\beta}$ 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 (1) with updates

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

(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 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 (1) 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. et al. (1998). *Independent inquiry into inequalities in health report*. The Stationery Office.

Amin-Chowdhury, Z. and S. N. Ladhani (2021). Causation or confounding: why controls are critical for characterizing long COVID. *Nature Medicine 27*(7), 1129–1130.

Appelman, B., B. T. Charlton, R. P. Goulding, T. J. Kerkhoff, E. A. Breedveld, W. Noort, C. Offringa, F. W. Bloemers, M. van Weeghel, B. V. Schomakers, et al. (2024). Muscle abnormalities worsen after post-exertional malaise in long COVID. *Nature communications 15*(1), 1–15.

Atchison, C. J., B. Davies, E. Cooper, A. Lound, M. Whitaker, A. Hampshire, A. Azo, C. A. Donnelly, M. Chadeau-Hyam, G. S. Cooke, H. Ward, and P. Elliott (2023). Long-term health impacts of COVID-19 among 242,712 adults in England. *Nature Communications 14*, 6588.

Ayoubkhani, D., K. Khunti, V. Nafilyan, T. Maddox, B. Humberstone, I. Diamond, and A. Banerjee (2021). Post-covid syndrome in individuals admitted to hospital with covid-19: retrospective cohort study. *BMJ 372*, n693.

Ayoubkhani, D., P. Pawelek, and C. Gaughan (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., S. K. van Zon, T. C. Olde Hartman, and J. G. Rosmalen (2022). Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study. *The Lancet 400*(10350), 452–461.

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

Behnood, S. A., R. Shafran, S. Bennett, A. X. Zhang, L. L. O'Mahoney, T. J. Stephenson, S. Ladhani, B. L. De Stavola, R. Viner, and O. V. Swann (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.

Birrell, P., J. Blake, E. Van Leeuwen, N. Gent, and D. De Angelis (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.

Boccia, S., W. Ricciardi, and J. P. Ioannidis (2020). What other countries can learn from Italy during the COVID-19 pandemic. *JAMA internal medicine 180*(7), 927–928.

Brainard, J., N. R. Jones, I. R. Lake, L. Hooper, and P. R. Hunter (2020). Community use of face masks and similar barriers to prevent respiratory illness such as COVID-19: a rapid scoping review. *Eurosurveillance 25*(49), 2000725.

Brainard, J., I. R. Lake, R. A. Morbey, N. R. Jones, A. J. Elliot, and P. R. Hunter (2023). Comparison of surveillance systems for monitoring COVID-19 in England: a retrospective observational study. *The Lancet Public Health*.

Brown, M., J. Gerrard, L. M. L, et al. (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.

Case, A. and A. Deaton (2017). Mortality and morbidity in the 21st century. *Brookings papers on economic activity 2017*, 397.

Chin, V., J. P. Ioannidis, M. A. Tanner, and S. Cripps (2021). Effect Estimates of COVID-19 Non-Pharmaceutical Interventions are Non-Robust and Highly Model-Dependent. *Journal of Clinical Epidemiology 136*, 96–132.

Ciment, J. (1999). Life expectancy of Russian men falls to 58. *British Medical Journal 319*, 468.

Cowling, B. J. and G. M. Leung (2020). Face masks and COVID-19: Don't let perfect be the enemy of good. *Eurosurveillance 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.

Ferguson, N. M., D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L. C. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, and A. C. Ghani. (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., S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature 584*(7820), 257–261.

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

Gray, A. M. (1982). Inequalities in health. The Black Report: a summary and comment. *International Journal of Health Services 12*(3), 349–380.

Gupta, S. (2001). Avoiding ambiguity. *Nature 412*, 589.

Hall, R. E. (2009). *Inflation: causes and effects*. University of Chicago Press.

Hanlon, P., F. Chadwick, A. Shah, R. Wood, J. Minton, G. McCartney, C. Fischbacher, F. S. Mair, D. Husmeier, and J. Matthiopoulos (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. and V. Prasad (2023). Comparability of Control and Comparison Groups in Studies Assessing Long COVID. *The American Journal of Medicine*, 10.1016/j.amjmed.2023.01.005.

Hastie, C. E., D. J. Lowe, A. McAuley, A. J. Winter, N. L. Mills, C. Black, J. T. Scott, C. A. O'Donnell, D. N. Blane, S. Browne, et al. (2022). Outcomes among confirmed cases and a matched comparison group in the Long-COVID in Scotland study. *Nature communications 13*(1), 5663.

Herridge, M. S. (2011). Recovery and long-term outcome in acute respiratory distress syndrome. *Critical care clinics 27*(3), 685–704.

Høeg, T. B., S. Ladhani, and V. Prasad (2023). How methodological pitfalls have created widespread misunderstanding about long COVID. *BMJ Evidence-Based Medicine*, 10.113bmjebm–2023–112338.

Holi, M., P. Sammallahti, and V. Aalberg (1998). A Finnish validation study of the SCL-90. *Acta Psychiatrica Scandinavica 97*(1), 42–46.

Holleyman, R. J., S. Barnard, C. Bauer-Staeb, A. Hughes, S. Dunn, S. Fox, J. N. Newton, J. Fitzpatrick, Z. Waller, D. J. Deehan, et al. (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.

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

Hopkins, R. O., L. K. Weaver, D. Pope, J. F. Orme Jr, E. D. Bigler, and V. Larson-Lohr (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.

Huang, C., Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet 395*(10223), 497–506.

Hutton, G. and A. Powell (2021, 10). 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., D. A. Jdanov, V. M. Shkolnikov, K. Khunti, I. Kawachi, M. White, S. Lewington, and B. Lacey (2021). Effects of covid-19 pandemic on life expectancy and premature mortality in 2020: time series analysis in 37 countries. *BMJ 375*.

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

Keeling, M. J., E. Brooks-Pollock, R. Challen, L. Danon, L. Dyson, J. R. Gog, L. G. Rincón, E. M. Hill, L. Pellis, J. M. Read, and M. J. Tildesley (2021). Short-term Projections based on Early Omicron Variant Dynamics in England. *medRxiv*.

Knock, E. S., L. K. Whittles, J. A. Lees, P. N. Perez Guzman, R. Verity, R. G. Fitzjohn, K. A. M. Gaythorpe, N. Imai, W. Hinsley, L. C. Okell, A. Rosello, N. Kantas, C. E. Walters, S. Bhatia, O. J. Watson, C. Whittaker, L. Cattarino, A. Boonyasiri, B. A. Djaafara, K. Fraser, H. Fu, H. Wang, X. Xi, C. A. Donnelly, E. Jauneijaite, D. J. Laydon, P. J. White, A. C. Ghani, N. M. Ferguson, A. Cori, and M. Baguelin (2020). Report 41: The 2020 SARS-CoV-2 epidemic in England: key epidemiological drivers and impact of interventions. *Imperial College London*.

Knock, E. S., L. K. Whittles, J. A. Lees, P. N. Perez Guzman, R. Verity, R. G. Fitzjohn, K. A. M. Gaythorpe, N. Imai, W. Hinsley, L. C. Okell, A. Rosello, N. Kantas, C. E. Walters, S. Bhatia, O. J. Watson, C. Whittaker, L. Cattarino, A. Boonyasiri, B. A. Djaafara, K. Fraser, H. Fu, H. Wang, X. Xi, C. A. Donnelly, E. Jauneijaite, D. J. Laydon, P. J. White, A. C. Ghani, N. M. Ferguson, A. Cori, and M. Baguelin (2021). Key epidemiological drivers and impact of interventions in the 2020 SARS-CoV-2 epidemic in England. *Science Translational Medicine 13*(602), eabg4262.

Linton, N. M., T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, and H. Nishiura (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.

Liu, Y., C. Morgenstern, J. Kelly, R. Lowe, and M. Jit (2021). The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC medicine 19*(1), 1–12.

Major, L. E., A. Eyles, and S. Machin (2021). Learning Loss since Lockdown: Variation across the Home Nations. COVID-19 Analysis Series. No. 023. *Centre for Economic Performance*.

Marmot, M., J. Allen, T. Boyce, P. Goldblatt, and J. Morrison (2020). *Health Equity in England: The Marmot Review 10 Years On*. The Health Foundation.

Matta, J., E. Wiernik, O. Robineau, F. Carrat, M. Touvier, G. Severi, X. de Lamballerie, H. Blanché, J.-F. Deleuze, C. Gouraud, 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.

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

McKeigue, P. M., D. McAllister, D. Caldwell, C. Gribben, J. Bishop, S. J. McGurnaghan, M. Armstrong, J. Delvaux, S. Colville, S. Hutchinson, et al. (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.

Mitze, T., R. Kosfeld, J. Rode, and K. Wälde (2020). Face masks considerably reduce COVID-19 cases in Germany. *Proceedings of the National Academy of Sciences 117*(51), 32293–32301.

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.

O' Mahoney, L. L., A. Routen, C. Gillies, W. Ekezie, A. Welford, A. Zhang, U. Karamchandani, N. Simms-Williams, S. Cassambai, A. Ardavani, et al. (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*, 101762.

O'Connor, R. C., K. Wetherall, S. Cleare, H. McClelland, A. J. Melson, C. L. Niedzwiedz, R. E. O'Carroll, D. B. O'Connor, S. Platt, E. Scowcroft, et al. (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.

ONS (2022). Self-reported long COVID and labour market outcomes, UK: 2022.

ONS (2023). Rising ill-health and economic inactivity because of long-term sickness, UK: 2019 to 2023.

Pamuk, E. R. (1985). Social class inequality in mortality from 1921 to 1972 in England and Wales. *Population studies 39*(1), 17–31.

Pritchard, M., E. A. Dankwa, M. Hall, J. K. Baillie, G. Carson, A. Docherty, C. A. Donnelly, J. Dunning, C. Fraser, H. Hardwick, et al. (2020). ISARIC clinical data report 4 October 2020. *medRxiv*.

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

Richardson, E., G. McCartney, M. Taulbut, M. Douglas, and N. Craig (2023). Population mortality impacts of the rising cost of living in Scotland: scenario modelling study. *BMJ public health 1*(1), e000097.

Riera, R., Â. M. Bagattini, R. L. Pacheco, D. V. Pachito, F. Roitberg, and A. Ilbawi (2021). Delays and disruptions in cancer health care due to COVID-19 pandemic: systematic review. *JCO global oncology 7*(1), 311–323.

Rougier, J., R. S. J. Sparks, K. V. Cashman, and S. K. Brown (2018). The global magnitude–frequency relationship for large explosive volcanic eruptions. *Earth and Planetary Science Letters 482*, 621–629.

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., L. B. Havdal, M. Drevvatne, E. M. Brodwall, L. L. Berven, T. Stiansen-Sonerud, G. Einvik, T. M. Leegaard, T. Tjade, A. E. Michelsen, et al. (2023). Prevalence and Characteristics Associated With Post–COVID-19 Condition Among Nonhospitalized Adolescents and Young Adults. *JAMA Network Open 6*(3), e235763–e235763.

Shine, C. (2022, April). 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. 19th December 2021. Technical report.

Stuckler, D., S. Basu, and M. McKee (2010). Budget crises, health, and social welfare programmes. *BMJ 340*.

Tkachenko, A. V., S. Maslov, A. Elbanna, G. N. Wong, Z. J. Weiner, and N. Goldenfeld (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.

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

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

Viboud, C., R. F. Grais, B. A. Lafont, M. A. Miller, and L. Simonsen (2005). Multinational impact of the 1968 Hong Kong influenza pandemic: evidence for a smoldering pandemic. *The Journal of infectious diseases 192*(2), 233–248. Table 1 gives excess mortality at 120 per 100000. Equivalent for Covid is around 130.

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

Wang, D., B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, et al. (2020). Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *Jama 323*(11), 1061–1069.

Ward, H., G. Cooke, M. Whitaker, R. Redd, O. Eales, J. C. Brown, K. Collet, E. Cooper, A. Daunt, K. Jones, et al. (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*.

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

Wood, S. N. (2020). Did COVID-19 infections decline before UK lockdown? *arXiv preprint 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.

Wood, S. N. and M. Fasiolo (2017). A generalized Fellner–Schall method for smoothing parameter optimization with application to Tweedie location, scale and shape models. *Biometrics 73*(4), 1071–1081.

Wood, S. N. and E. C. Wit (2021). Was R<1 before the English lockdowns? On modelling mechanistic detail, causality and inference about Covid-19. *PLoS ONE 16*(9), e0257455.

Wood, S. N., E. C. Wit, M. Fasiolo, and P. J. Green (2020). COVID-19 and the difficulty of inferring epidemiological parameters from clinical data. *The Lancet Infectious Diseases*.

Wu, J. T., K. Leung, M. Bushman, N. Kishore, R. Niehus, P. M. de Salazar, B. J. Cowling, M. Lipsitch, and G. M. Leung (2020). Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nature Medicine 26*(4), 506–510.

Zhou, F., T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, et al. (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.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. Profile Books.