Measuring and controlling the Covid-19 pandemic

Philip Thomas*

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

A system has been devised to measure the social distancing index (SDI)—also known as the basic reproduction number—on a daily basis, as well as the R-rate and the number of active infections. The predictor–corrector measurement tool, which incorporates both a single cohort and a two-cohort model of the Covid-19 epidemic, has been applied to England. The measurements have been corroborated against spot figures calculated from data for England from the Office of National Statistics Coronavirus Infection Survey Pilot. They give comparable accuracy and are available 10 days earlier. Accurate measurement has required the development of a correction factor to be applied to the Cases by Date Reported based on the number of tests carried out. Proper comparability is now possible for the first time between daily numbers of cases on different dates. It has been necessary, in addition, to account for the level of prior immunity in the population due to T-cells. Recent research showing that roughly 1 in 8 people carries such immunity is confirmed by the validation of the predictor–corrector model against ONS survey data on active infections. The measurement system has shown that England's R-rate was either approaching or below 1.0 at the time the country entered its second lockdown on 5 November 2020. The action of the government in imposing the lockdown suggests it might not have been aware of this situation. The capability of the new measurement system to provide timely feedback to decision-makers allows them the freedom for the first time to control the SDI to a setpoint of their choice. Policymakers need no longer see their choices of Covid-19 strategies as binary: either lock down or let the virus rip. The predictor–corrector measurement system allows a third way, whereby the SDI is controlled to a series of setpoints chosen to minimize the overall harm to the nation's health and economy. The point is illustrated with two scenarios that consider the effects of mass vaccination of the English population from December 2020 to the end of the epidemic in late 2021.

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

Nanotechnology Perceptions **16** (2020) 267–330
Received 13 October 2020; *accepted* 30 November 2020

doi: 10.4024/N15TH20A.ntp.16.03

The R-rate is kept below 1.0 at all times in the first scenario, whereas in Scenario 2 the SDI is set initially at the value, 1.5, it had just before England went into the first set of Tiers on 17 October 2020. However, the R-rate quickly falls below 1.0 because of the immunity from various sources: from T-cells, from Covid-19 recovery and from vaccination. The total number of deaths is predicted to be similar under both scenarios, but the second allows somewhat greater economic freedom at the start.

Keywords: basic reproduction number, coronavirus, Covid-19, epidemic, exit strategy, lockdown, measurement, R-rate, social distancing index, vaccination

1. Introduction

Originating in Wuhan, China in late 2019, Covid-19 developed swiftly into a pandemic affecting all regions of the world. It is estimated that 34.5 million people worldwide had developed the new respiratory disease by the beginning of October 2020. Most recovered but over a million people are thought to have died.[1] The causative agent,[2] coronavirus SARS-CoV-2, induces a remarkably wide variation in response in those contracting it. About half have no symptoms and a further 30–40% experience relatively mild effects;[3] it is fatal for a minority of people.

Governments imposed restrictions on people's mixing in attempts to control the spread of the virus during the first wave, March to April 2020. The new concept of "social distancing" (also called "physical distancing") came into being: people were urged to stay away from others so as not to infect or be infected. Realizing the economic and social harm of harsh restrictions, governments then relaxed the constraints as the number of cases fell in the summer of 2020, only to reimpose them later when there was a second wave of infection, September to November 2020. The process seemed haphazard at times; people questioned why some activities were allowed while others, which seemed to pose no more risk, were prohibited. There appeared to be no dependable yardstick to judge either the effect individual measures were having or how well people as whole were social distancing.

But a very good measure already exists of how well a population is keeping to the rules for social distancing, namely the *basic reproduction number*, R_0, which tells us how many people in a fully "naïve" population will be infected by an average person carrying the infection before that person recovers. It starts out at a high value (about 3 in the UK in March 2020) but will vary with time as people adapt their behaviour in response to infection levels. It will decrease when the government brings in restrictions and it will increase when those restrictions are

1 World Health Organization, WHO Coronavirus Disease (COVID-19) Dashboard (3 October 2020) https://covid19.who.int/?gclid=EAIaIQobChMI5dP_v52Z7AIVw7HtCh2oZAeuEAAYASAAEgIW-_D_BwE

2 Covid-19 is caused by coronavirus of the type SARS-CoV-2 (Severe Acute Respiratory Syndrome-Coronavirus-2). Except in §5, which mentions common-cold coronaviruses, only one coronavirus is considered in this paper, the convention followed by the Office of National Statistics and UK national newspapers will be adopted, meaning that "Covid-19" and "coronavirus" will be treated as synonymous.

3 The asymptomatic fraction may be even higher. See Petersen, I. and Phillips, A., Three quarters of people with SARS-CoV-2 infection are asymptomatic: analysis of English household survey data. *Clinical Epidemiology* (9 October 2020) file:///C:/Users/Philip/Downloads/clep-276825-three-quarters-of-people-with-sars-cov-2-infection-are-asymp.pdf

relaxed. R_0 is thus an ideal measure of how well people are socially (physically) distancing. Its importance in this regard merits its renaming as the social distancing index (SDI), a practice introduced and followed in this paper. A low value of the SDI means that people are adopting social distancing wholeheartedly; a high value means that not much social distancing is being practised. It is a measure that governments can focus on to get control of the pandemic.

The R-rate, R, has received a lot of attention from the UK government. But while it can be a useful indicator, it is a poor measure of how well people as a whole are conforming with social distancing restrictions. This is because it conflates two separate factors:

(i) the extent to which people are observing social distancing; and
(ii) the fraction of people in the population who have contracted the disease in the past and are thus immune.

All susceptible individuals have a large measure of direct control of their social distancing, but they have no control of how many people have had the disease previously. This is something that relates to the past actions of others (not themselves, because they have, by definition, not had the disease), which actions cannot now be changed.

Reliance on the R-rate to assess how well people have been social distancing is fraught with problems. As an example, it is shown in §12.1 how people can do less and less social distancing but still keep the R-rate at a low level when the number of susceptible people is falling. Moreover, designing a control scheme around a variable over which control is only partial is a poor strategy when there is an alternative variable, the SDI, that can be measured and presented for all to see and tailor their actions accordingly.

A system has been devised to measure SDI on a daily basis. The system measures the number of active infections *and* the R-rate. It offers timely feedback for the first time to decision-makers, allowing them to control SDI to a setpoint of their choice. This paper will describe the measurement system and how it works. It will be applied to England.

The paper is arranged as follows. §2 provides background on the Covid-19 pandemic, the level of individual risk, the wide range of responses to the infection by governments, the measures adopted by the United Kingdom in trying to contain the crisis and the economic consequences that have followed. Attention is given to the succinct statement by *The Times* columnist Daniel Finkelstein of the binary dilemma that the UK government appears to think it faces. General requirements for the control of any process are set out.

§3 considers in detail the properties of the variable that provides the input to the measuring system, namely Cases by Date Reported. The speed advantage of using this parameter will be discussed. §4 explains how to correct for the widely varying numbers of viral tests administered each day. §5 considers the important topic of T-cell immunity, and how this has now been quantified.

§6 contains mathematical descriptions of the predictor, which is based on a single-cohort representation of England's population, and the corrector, which uses a two-cohort model. §7 shows how the predictor–corrector system has been validated against independent empirical data from Office of National Statistics (ONS) surveys to find the number of active infections in England.

§8 shows how the R-rate may be deduced using data from ONS weekly Covid-19 infection surveys and §9 compares the R-rates measured by the predictor–corrector with those based on the ONS survey.

§10 shows how the SDI may be estimated from ONS weekly survey data, while §11 provides a comparison of the SDI values produced by the predictor–corrector with those based on ONS survey data.

§12 describes two scenarios starting in early December 2020 and continuing until the epidemic is brought to an end in England in late 2021, assisted by mass vaccination programmes. Different time profiles of the SDI will be used.

§13 contains the discussion and §14 the conclusions.

There are six appendices. Appendix A describes the development of the Temporary Age Burden (TAB) method for assessing any individual's risk from catching Covid-19. Appendix B explains the corrections made to the data for the computer error for England, 25 September 2020 to 4 October 2020.

Appendix C describes how the expected value and variance may be found for the R-rate estimated from ONS surveys. Appendix D explains how the effect of differences in the number of tests on different days can be corrected for.

Appendix E describes how the expected value and variance may be found for the social distancing index estimated from ONS surveys.

Appendix F explains the epidemiological modelling of vaccination, while Appendix G outlines the method used to model the preferential inoculation of the most vulnerable.

2. Background

2.1 The risk of dying from Covid-19

The substantial fraction of infected people experiencing either mild or no symptoms means that incidence will never be fully reported. With the total number of sufferers uncertain, the true infection fatality rate cannot be known with total precision, now or in the future. The range of expert opinion is said to lie between 0.5% and 1%.[4]

The severity of the disease symptoms depends strongly on age. Covid-19 deaths are concentrated in the elderly, with 89.3% of coronavirus deaths occurring in those over 65 in England and Wales for the period from December 2019 to 18 September 2020[5] (see Figure 1). Moreover, the majority of Covid-19 deaths are associated with serious conditions being present already, for example dementia and Alzheimer's disease, ischaemic heart disease, influenza and pneumonia, chronic lower respiratory disease, diabetes. 91.1% of the people dying from Covid-19 in England and Wales between March 2020 and the end of June 2020 were found to be suffering from at least one such pre-existing condition.[6] It is reported that

---

[4] Beasley, D., We may never know COVID-19's real death rate. Here's why. World Economic Forum (29 September 2020) https://www.weforum.org/agenda/2020/09/death-rate-fatality-covid-19-coronavirus-disease-pandemic-science

[5] Office of National Statistics, Deaths registered weekly in England and Wales, provisional: week ending 18 September 2020 (29 September 2020) https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsregisteredweeklyinenglandandwalesprovisional/weekending18september2020

[6] Office of National Statistics, Deaths involving COVID-19, England and Wales: deaths occurring in June 2020 (17 July 2020) https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingcovid19englandandwales/deathsoccurringinjune2020#characteristics-of-those-dying-from-covid-19

fewer than 400 UK citizens under the age of 60 without comorbidities had died from the coronavirus by 11 October 2020.[7]

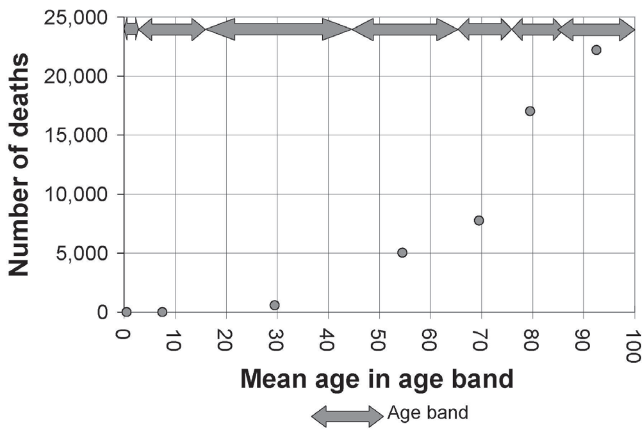

Figure 1. Covid-19 deaths by age band, England and Wales to 18 September 2020. Age bands: < 1, 1 – 14, 15 – 44, 45 – 64, 65 – 74, 75 – 84, > 85.

The greater impact on the aged is evident from Figure 1, but the smaller number of people in the older age groups means that their relative vulnerability to the disease is higher than might be supposed from a first glance at the graph. The Covid-19 mortality rate (per year per 100,000) in England, shown in Figure 2[8] for April 2020, the peak month for the first wave of infections, is found to fall away very rapidly with decreasing age.

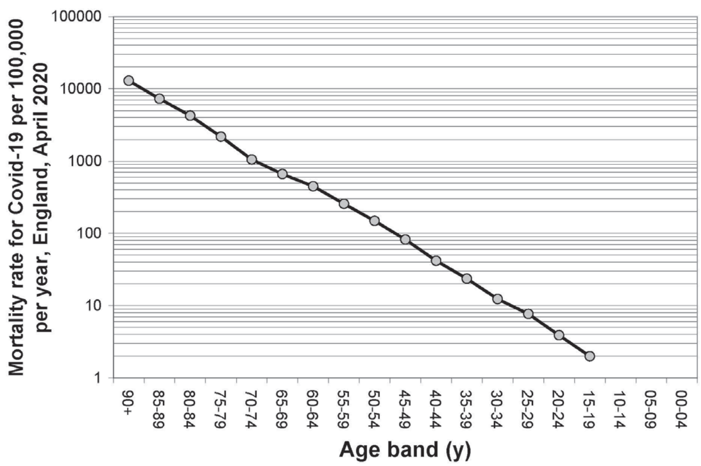

Figure 2. Covid-19 mortality rate per year per 100,000 in the different age bands, based on deaths recorded in England during April 2020.

[7] Livermore, D., Protect elderly, but let's see life back to normal for low-risk. *Sunday Telegraph* (11 October 2020) https://digitaleditions.telegraph.co.uk/data/378/reader/reader.html?#!preferred/0/package/378/pub/378/page/39/article/90592

[8] Office of National Statistics, Deaths involving COVID-19, England and Wales: deaths occurring in

Such mortality rates may be related to an age-dependent infection fatality rate for Covid-19 through assessing the number of people likely to have been infected in each age band. Spiegelhalter[9] has made the important observation that the relationship between the Covid-19 infection fatality rate and age is similar to the age dependence of the normal mortality rate. Figure 3 shows an estimate of the Covid-19 infection fatality rate[10] compared against the UK's normal mortality rate per year.[11]

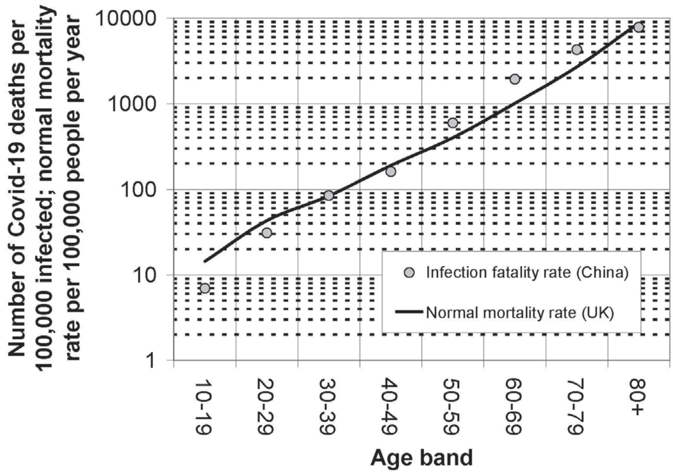

Figure 3. Comparison of fatality rate after infection with Covid-19 against normal mortality rate per year, across the age bands.

The two rates are roughly equal. To a first approximation, the probability of a dose of Covid-19 killing the just-infected person aged 20 or older is the same, on average, as the chance of a person of that age dying from normal, non-Covid causes over the next 12 months. Hence his or her normal chance of dying over the coming year is doubled. This result holds for all ages from 20 to 100, even though the risk of death increases more than a thousandfold over the range in each case. Meanwhile anyone contracting Covid-19 when below the age of 20 will face a very low risk of dying, as can be judged from Figure 1.

Spiegelhalter's insight can be re-expressed in terms of the increase in age one would need to experience before the extra risk of living normally but at a greater age over the next 12 months would match one's risk of dying from Covid-19 having recently been infected. So, for

April 2020 (15 May 2020) https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsinvolvingcovid19englandandwales/deathsoccurringinapril2020

9 Spiegelhalter, D., Use of 'normal' risk to improve understanding of dangers of covid-19. *BMJ* **370** (2020) m3259 (published 9 September—see http://dx.doi.org/10.1136/bmj.m3259).

10 Verity, R. et al., Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infectious Diseases* **20** (2020) 669–677 (https://doi.org/10.1016/S1473-3099(20)30243-7).

11 Office of National Statistics, National Life Tables: UK (24 September 2020) https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/datasets/nationallifetablesunitedkingdomreferencetables

someone aged between 20 and 59, the risk of dying from Covid-19 having just caught it equals the extra risk of living normally and Covid-free for the next year but being 9 years older throughout the period. The extra age is then removed entirely when the year is up. This temporary age burden (TAB) is 8 years for someone between 60 and 70. Meanwhile those catching Covid-19 over the age of 70 will need to add on only 6 years to their age for a year, reflecting the fact that the increase in the normal hazard rate with age has speeded up (see Appendix A for a derivation of the TAB method for expressing the risk of Covid-19). Casting the risk of Covid-19 once contracted in terms of living with an extra age burden for 12 months draws on people's experience and our knowledge that older people have a higher chance of dying from natural causes. It circumvents, to a large extent, the problem Spiegelhalter noted with his original formulation, namely that it applied only *on average* to a person of a given age. Under the TAB method, each individual can consider how he or she personally is likely to be faring in some years' time.

It would be reasonable for healthy people to see themselves as part of a cohort likely to continue in good health for a good many years to come. This would cause them to weigh up the effect of the temporary age burden optimistically. Meanwhile those in poorer health might see themselves as part of a struggling cohort and come to a more pessimistic conclusion.

Take, for example, a 65 year-old catching Covid-19. He or she will face the same risk from the coronavirus as the additional risk he/she would take on by having to live Covid-free but as that person's 73 year-old self for the next 12 months. The absolute value of the additional risk imposed by the extra years will be comparatively low for a fit and healthy person. The risk burden will, however, be much higher if the individual is frail and suffering from a serious underlying condition. By considering his or her specific health situation in the way outlined, each person is likely to come up with an interpretation of the message that will be broadly correct.

Clearly, catching Covid-19 increases the normal risk of dying, especially for those over the age of 65. But the demystification provided by the TAB method offers everyone a way of gaining a personal understanding of his or her Covid-19 risk. Spiegelhalter's original formulation shows that risk of death from contracting Covid-19 is comparable with the other normal risks faced in everyday life. The TAB method shows that it is comparatively low in absolute terms for most people, with the notable exceptions of those whose health is already compromised and the old, with the vulnerability of older people being much greater if they are also in fragile health.

2.2 The responses of governments and economic effects

Among government responses to the pandemic, Sweden's stands out as unusual because no universal lockdown was imposed during the first wave of coronavirus. Schools for children under 16 stayed open with small class sizes, hygiene measures and social distancing. Learning moved online for 3 months from 17 March 2020 for children of 16 and over and for university students, but face-to-face teaching resumed from the middle of June 2020.[12] Gatherings of more

[12] Iacobucci, G., Covid-19: Comparing Sweden's response with the UK's is "misleading," experts argue. *BMJ* **370** (2020) m3765 (published 28 September—see http://dx.doi.org/10.1136/bmj.m3765). The article reports on a SAGE paper and media briefing that expressed "concern that the Swedish model may be given some credence".

than 50 people were forbidden. Visiting elderly care homes was banned until 1 October 2020.[13] Travel from Non-EU/EEA countries was prohibited up to 31 October 2020.[14] But, taken as a whole, Sweden's public health measures were significantly less restrictive than those imposed by other countries in Europe.

Daily cases of Covid-19 in Sweden peaked at just over 1000 at the end of June 2020 (based on a 7-day moving average) before falling to around 300 per day by the second week of July 2020. New daily cases continued at this level until the beginning of September 2020, when they began to rise, reaching just under 500 per day by the last week of September 2020. There had been 5895 coronavirus deaths recorded in Sweden by 5 October 2020.[15] Given Sweden's 2020 population of 10.1 million (88% urban),[16] this implies a Covid-19 death rate of 584 per million by that time.

The economic effect of the coronavirus epidemic on Sweden has been considerable. Gross domestic product (GDP) fell by 8.6% in the second quarter of 2020, recording its largest single quarterly drop in modern history. This was, however, significantly less that the fall of 12.2% in the same period across the Eurozone.

England and the rest of the UK conformed to the pattern adopted by most countries across the world and imposed a strict lockdown. This came into effect on the evening of 23 March 2020, when the hospitality and entertainment industries were instructed to close. Schools were asked to suspend teaching soon after and universities followed suit. Shops deemed to be nonessential (grocery supermarkets and pharmacies were exceptions) were told to close. People were instructed to stay in their own dwellings except for work, medical reasons, buying food or taking exercise once a day, locally and for no longer than an hour. Employees in all industries were expected to work at home if at all possible. Visits to second homes and overnight stays in general were forbidden except in a limited set of circumstances. Construction was permitted by the Government, but some regional authorities were hostile to the idea and many construction firms closed. A large number of other businesses where home working was not feasible decided to suspend all or most of their activity. Effectively about a third of the UK's economy ceased trading at the end of March 2020.

The lockdown, as originally announced, was expected to last 3 weeks, subject to review. But it was extended at the end of the first period for a further three weeks and later for longer still. The easing of restrictions started 7 weeks after they had been introduced, with people in all industries except hospitality, entertainment and general retail being encouraged to return to work on 13 May 2020. That date also marked the lifting of restrictions on travel away from the home in England. Primary school children in Year 1 and Year 6 went back to school on 1 June 2020.

Shopping centres and high streets in England were able to welcome customers again on 15 June 2020, subject to social distancing measures. Hospitality was allowed to reopen on 4 July

---

[13] Emergency information from Swedish Authorities: International travel restrictions https://www.krisinformation.se/en/hazards-and-risks/disasters-and-incidents/2020/official-information-on-the-new-coronavirus/travel-restrictions (accessed 5 October 2020).

[14] Sweden to lift ban on visits to elderly care homes. *The Local* (15 September 2020) https://www.thelocal.se/20200915/sweden-to-lift-ban-on-visits-to-elderly-care-homes

[15] Sweden coronavirus cases https://www.worldometers.info/coronavirus/country/sweden/ (accessed 5 October 2020).

[16] Sweden population https://www.worldometers.info/world-population/sweden-population/ (accessed 5 October 2020).

2020 in England[17] and hairdressers from 13 July 2020, the latter subject to customers wearing a mask. Masks became mandatory in shops in England on 24 July 2020 (although not in Wales until 11 September 2020).

A quarantine period of 14 days for anyone travelling back from overseas was introduced in England on 8 June 2020, but this restriction was eased on 3 July 2020, when it was announced that return from 73 countries would no longer be subject to self-isolation. Within 4 weeks, however, the number of eligible countries began to be reduced, starting with the delisting, on 26 July 2020, of a favourite UK holiday destination, Spain. Further removals from the list were to follow, including another popular tourist destination, France, on 13 August 2020.[18]

Additional restrictions on social contact were imposed on Leicester as early as 29 June 2020, and further "local lockdowns" were introduced during the summer. As of 5 October 2020, a quarter of England's population was facing some sort of extra restriction, three quarters in Wales, a third in Scotland and 8% in Northern Ireland.[19] A "rule of 6" was introduced in England on 14 September 2020, whereby social gatherings of more than 6 people became illegal. Moreover, people were at the same time discouraged once again from going to work if working from home was possible.

Daily cases of Covid-19 in the UK reached a peak of about 5,000 in the middle of April 2020, based on a 7-day average, before falling to around 550 per day by the end of the first week of July 2020. New daily cases continued at this low level until the beginning of September 2020, when they began to rise, reaching roughly 8,000 per day by the end of the month and reached 30,000 per day by late October 2020.[20]

The number of polymerase chain reaction (PCR) swab tests for active Covid-19 infections that were processed daily rose from 22,000 to more than a quarter of a million between the middle of April 2020 and late September 2020, as shown in Figure 4. This tenfold increase in testing introduces a confounding factor into the comparability of daily case numbers over longer time periods.

By 5 October 2020, 42,628 deaths attributed to coronavirus had been recorded in the UK. With a population in 2020 of 67.9 million (83% urban), the UK's Covid-19 death rate to that date was 628 per million, slightly higher than Sweden's.

The effect of the coronavirus epidemic, including the UK's prolonged lockdown, caused an exceptionally severe recession, with GDP falling by 19.8% in Quarter 2 of 2020. This represented the largest quarterly contraction in the UK economy since records began in 1955 and followed a fall of 2.5% in the previous 3 months. UK output between April and June 2020 was 21.5% lower than in the same period in 2019.

The Chancellor of the Exchequer announced the Coronavirus Job Retention Scheme, known as the "furlough scheme" for short, on 20 March 2020, whereby employers could seek

[17] Wales, Scotland and Northern Ireland often waited a week or so before introducing the same or similar easing.

[18] Press Association Reporters, Coronavirus: Timeline of key events since UK was put into lockdown six months ago. *The Independent* (23 September 2020) https://www.independent.co.uk/news/uk/home-news/coronavirus-uk-timeline-lockdown-boris-johnson-pubs-test-and-trace-vaccine-b547630.html

[19] Covid rules: How much of the UK is now under some sort of lockdown? (2 October 2020) https://www.bbc.co.uk/news/uk-england-52934822

[20] UK coronavirus cases https://www.worldometers.info/coronavirus/country/uk/ (accessed 6 October 2020).

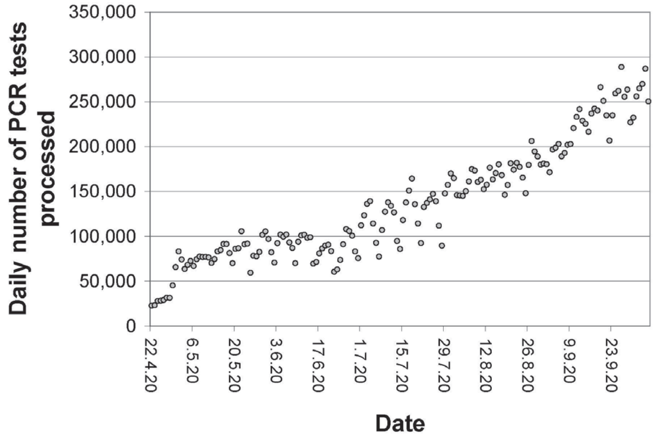

Figure 4. Covid-19 PCR swab tests processed in the UK each day.

reimbursement of up to 80% of a furloughed worker's wages, subject to a cap of £2,500 per month per furloughed employee.[21] Originally planned to cover the period 1 March–31 May 2020, it was subsequently extended. Employers needed to pay National Insurance and pension contributions from 1 August 2020, and the 80% government contribution was reduced to 70% in September 2020 and 60% in October 2020, its final month.[22] 9.6 million jobs were supported under the furlough scheme.[23] A new scheme was announced by the Chancellor on 9 October 2020 ahead of new Government restrictions that were expected to take the form of greater curbs for cities in the north of England from 14 October 2020. The new Job Support Scheme provided employees with two thirds of their salaries up to a limit of £2,100 per month if they work in businesses across the UK required to close their premises due to additional coronavirus measures.

UK public sector net debt at the end of July 2020 exceeded £2 trillion (2×10^{12}) for the first time, £0.2276 trillion more than at the same point the previous year. It represented 100.5% of GDP, an increase of 20.4 % on the year before. This was the first time that debt had risen above 100% of GDP since the financial year ending March 1961.[24] Public sector borrowing continued on a massive scale in August 2020: £35.9 billion (35.9×10^9) are estimated to have been borrowed during the month, 6½ times the figure for August 2019.[25]

---

21 Jeffcott, R.B., Green, G., Ashmore, D. and Heaton, A., Coronavirus Job Retention Scheme as at 23 March 2020. Reed Smith Employment Watch (26 March 2020) https://www.employmentlawwatch.com/2020/03/articles/employment-uk/coronavirus-job-retention-scheme-as-at-23-march-2020/#:~:text=The%20Scheme%20allows%20an%20employer,alternative%20to%20terminating%20their%20employment

22 HM Revenue and Customs, Changes to the Coronavirus Job Retention Scheme (1 July 2020) https://www.gov.uk/government/publications/changes-to-the-coronavirus-job-retention-scheme/changes-to-the-coronavirus-job-retention-scheme

23 Clark, D., Number of jobs furloughed under the job retention scheme in the UK 2020 (9 September 2020) https://www.statista.com/statistics/1116638/uk-number-of-people-on-furlough/

24 Office of National Statistics, Public sector finances, UK: July 2020 (21 August 2020) https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/publicsectorfinance/bulletins/publicsectorfinances/july2020

25 Office of National Statistics, Public sector finances, UK: August 2020 (25 September 2020) https://www.ons.gov.uk/economy/governmentpublicsectorandtaxes/publicsectorfinance/bulletins/publicsectorfinances/august2020#main-points

It has previously been pointed out that adopting Covid-19 restrictions that reduced GDP per head by 6.4% or more over a prolonged period would cost more life than would be restored.[26] This is because a nation's economy and its health are so strongly linked that at some point they become different aspects of the same thing. The effect is seen with children born in some poorer, sub-Saharan states, who will survive to a mean age of only 55, while those born in wealthier Japan live three decades longer on average.

A definite relationship exists between population-average life expectancy and GDP per head for all the world's nations—the "Bristol curve", which can be explained using the Judgement (J)-value[27]—see Figure 5, which plots the average life to come across all ages against GDP per head. Increases in GDP per head will increase life expectancy, as has been explained quantitatively for the UK between 1985 and 2005.[28] But, by the same token, reducing the nation's GDP per head will reduce its life expectancy, either in absolute terms or from the value it would have had. This latter effect was seen in the 2007–2009 recession in the UK, where the UK's steady growth in life expectancy stalled about 2 years after the drop of 6% in GDP per head and has not recovered to any great extent. A reduction in life expectancy means that more people die young, an effect that can be quantified as the number of years of life that will be lost across the nation.[29]

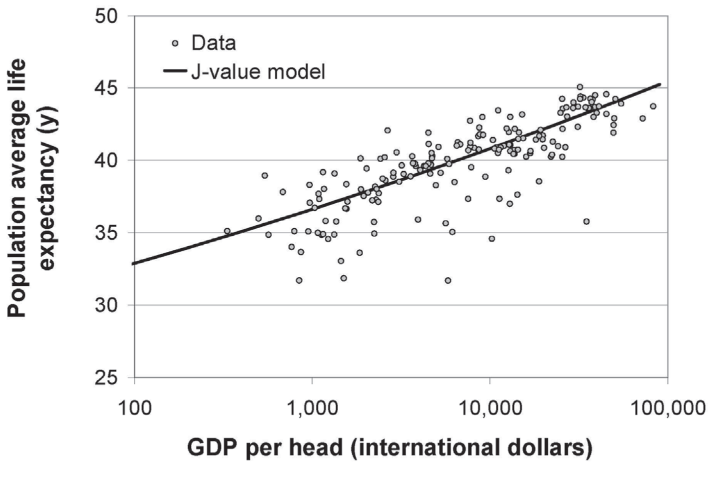

Figure 5. Bristol curve of population-average life expectancy v. GDP per head for 180 out of 193 countries in the UN, together with J-value prediction.

---

[26] Thomas, P., J-value assessment of how best to combat COVID-19. *Nanotechnology Perceptions* **16** (2020) 16–40 (http://www.colbas.org/ntp/opnAxs/N02TH20A.pdf).

[27] Thomas, P. and Waddington, I., Validating the J-value safety assessment tool against pan-national data. *Process Safety and Environmental Protection* **112A** (November 2020) 179–197 (https://ac.els-cdn.com/S0957582017302896/1-s2.0-S0957582017302896-main.pdf?_tid=e4b52f78-d2c5-11e7-b19e-00000aacb360&acdnat=1511713484_c1408cbb7d734eb48189c8bec89fcd68).

[28] Thomas, P., Corroboration of the J-value model for life-expectancy growth in industrialized countries. *Nanotechnology Perceptions* **13** (2017) 31–44.

[29] The loss of life due to the recessionary effects of lockdown has now been acknowledged by official sources: Department of Health and Social Care, Office for National Statistics, Government Actuary's

Addressing the NHS Providers Conference on 8 October 2020, the Secretary of State for Health, Matt Hancock, laid out his government's current coronavirus policy in clear terms: "Our strategy is simple – suppress the virus and support the economy, education and the NHS until a vaccine can make us safe."[30] Suppressing the virus means keeping the effective reproduction number or R-rate, R, below 1.0; R is the number of people the average person with an active infection will infect before recovering under current conditions; note that greater resistance to infection develops as a result of people contracting the disease and becoming immune on their recovery. $R < 1.0$ implies that the number of infected people falls inexorably until there are no new infections and the disease dies out entirely.

While the strategy outlined above has obvious superficial attractions, the problem is that most of the lockdown measures then need to be maintained or reimposed. In fact, the Prime Minister felt the need to reimpose a 4-week lockdown on England on 5 November 2020,[31] followed by a stricter Tier system on 2 December 2020, placing 99% of the country in the top two Tiers.[32]

2.3 The supposed binary choice

Writing a week after the rule of six was introduced on 14 September 2020, but before Matt Hancock offered the clarification above, Daniel Finkelstein, a columnist with *The Times*, realized that the Government's policy was contingent on obtaining a vaccine, and suggested that the nature of the decision could be no other than binary:[33]

> We must choose. We can allow the virus to spread, aiming to achieve herd immunity as quickly as possible. Or we can try to hold it off until we get a vaccine or treatment.

He inclined to the latter option and proposed that more restrictions would be sensible, all the while hoping that a vaccine would turn up fast enough, although he had qualms about how certain this was. His doubts on that point received independent support from a Royal Society report released a few days later,[34] with one of the authors, Prof. Nilay Shah, saying "Even when

---

Department and Home Office, *Direct and Indirect Impacts of COVID-19 on Excess Deaths and Morbidity* (15 July 2020), which concludes that "when morbidity is taken into account, the estimates for the health impacts from a lockdown and lockdown-induced recession are greater in terms of QALYs than the direct COVID-19 deaths".

30 Donnelly, L., Precisely how prepared is the NHS for what comes next? *Daily Telegraph* (9 October 2020) https://digitaleditions.telegraph.co.uk/data/377/reader/reader.html?#!preferred/0/package/377/pub/377/page/18/article/90337

31 Covid-19: PM announces four-week England lockdown (31 October 2020) https://www.bbc.co.uk/news/uk-54763956

32 Neilan, C. and Jones, A., Tiered system passes Commons but Boris Johnson suffers biggest backbench rebellion to date. *The Telegraph* (1 December 2020) https://www.telegraph.co.uk/politics/2020/12/01/covid-tier-system-vote-parliament-commons-brexit-news-latest/

33 Finkelstein, D., Boris Johnson has undercooked his Covid plan. An understandable desire to limit damage to the economy means that the PM's new restrictions don't go far enough. *The Times* (22 September 2020) https://www.thetimes.co.uk/article/boris-johnson-has-undercooked-his-covid-plan-mh0rbw0rb

34 Royal Society DELVE initiative, *SARS-CoV-2 Vaccine Development & Implementation; Scenarios, Options, Key Decisions* (1 October 2020) https://rs-delve.github.io/reports/2020/10/01/covid19-vaccination-report.html

the vaccine is available it doesn't mean within a month everyone is going to be vaccinated. We're talking about six months, nine months ... a year. There's not a question of life suddenly returning to normal in March."[35]

The epidemiological benefits of vaccinations will be examined quantitatively in the scenarios given in §12. But, by the J-value-based argument detailed previously,[26] a repeat of something like the original lockdown, even if maintained "only" to March 2021, might cause far too much indirect loss of life, outweighing any benefit it could bring to those vulnerable to Covid-19.

Fortunately, the choice need not be binary. Four options were considered in a previous paper.[36] The first and fourth options corresponded to Finkelstein's, and neither was optimal. But there were two strategies in between. The second option was to move out of lockdown as quickly as possible without overstraining the health services, while the third was to leave lockdown completely in 2020 while minimizing Covid-19 cases.

Option 2 would involve letting the basic reproduction number rise to 1.18, while under Option 3 R_0 would increase to 1.45. Both these R_0-values fall a long way short of allowing an uncontrolled outbreak or "letting the virus rip", as it is sometimes described.[37] But Option 3, which involved moving out of lockdown completely by the end of 2020, might have presented a greater challenge to the health services, even though it would have preserved more life in the long term through minimizing the damage to health caused by recession. Hence a recommendation[39] was made to the Treasury Committee that the best course of action was to *steer as close as possible to the rapid but controlled exit from lockdown exemplified by Option 3*.[40]

2.4 Measurement and control

Much emphasis has been placed during the coronavirus crisis, by politicians and the media, on measuring people's viral load and their antibody levels—"testing". But less attention has been given, in the UK at least, to *measuring* what appears to be the key determinant of government policy, the "R-rate". Lord Kelvin set out the universal merits of measurement in a quotation that has become justly famous:[41]

---

[35] Blakely, R., Life won't return to normal even after a vaccine, warn scientists. *The Times* (2 October 2020) https://www.thetimes.co.uk/article/life-wont-return-to-normal-even-after-a-vaccine-warn-scientists-rqqzm2bwf

[36] Thomas, P., The options for the UK leaving the coronavirus lockdown of 2020. *Nanotechnology Perceptions* **16** (2020) 130–150 (http://www.colbas.org/ntp/opnAxs/N08TH20A.pdf)

[37] In an echo of Daniel Finkelstein's argument that the choice was binary, the UK's Minister for Care, Helen Whately, used these words in the course of the House of Commons debate on Public Health on 6 October 2020: "The alternative—just allowing the virus to let rip—simply cannot be the right thing to do".[38]

[38] https://hansard.parliament.uk/commons/2020-10-06/debates/B97F9EF6-950B-4012-9568-6CAEAEC5E3FA/PublicHealth

[39] Thomas, P., EIC0563: Written evidence submitted by Philip Thomas, Professor of Risk Management, South-West Nuclear Hub, University of Bristol, Statement 496, Treasury Committee, Economic impact of coronavirus: the challenges of recovery (11 September 2020) https://committees.parliament.uk/writtenevidence/6369/html/

[40] The beneficial effect has been confirmed and quantified, as will be explained in §5. This improves the outlook for Option 3 as, of course, does vaccination.

[41] Lord Kelvin, Electrical units of measurement. In: *Popular Lectures and Addresses, vol. 1* (1889) (delivered 3 May 1883). See https://www.oxfordreference.com/view/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236

> When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.

In line with Lord Kelvin's ideas, control engineers from the 1930s onwards have seen the quantitative and timely measurement of a parameter as the indispensable prerequisite for effective control in any process, from jet fighters to nuclear reactors.

While the UK Government frequently stated its desire to "keep the R-rate below 1.0", it seemed content to quote, every so often, a wide range of possible values produced by different research groups in the UK. It began publishing updated official measurements of the parameter as a matter of routine only since October 2020. There seems to have been an assumption up to then that, since the policy objective was simply to keep R under 1.0 until a vaccine appeared, it was sufficient to observe whether cases were falling, in which case $R < 1$ and nothing needed to be done, or rising, implying $R > 1$ and the need to impose extra restrictions. But such a strategy is really only applicable during a full-scale lockdown, and it is crude even then. It will inevitably fail when restrictions are eased, driven by the objective of restarting the economy and then maintaining a reasonable level of prosperity. Decision-makers then need to know not just whether the R-rate is below 1.0 or not—it will not be—but how far it is above unity.

In fact, the more fundamental R_0 is a much more important figure than the R-rate R (given by equation 10 below). As described in the Introduction, R_0 measures the degree of social distancing that the population as a whole has adopted. For that reason, a more useful and meaningful description for R_0 is the "social distancing index" (SDI).

Calculating the SDI simply from a list of restrictions that are in force is unrealistic, as it is not possible to judge the effects of individual measures on the basic reproduction number R_0 with any accuracy. Fortunately an alternative and much better way exists that makes use of the number of new cases published daily on the UK government's Coronavirus Dashboard. These figures were problematic in the first months of the epidemic in the UK, but the reorganized and much improved system of testing and reporting the numbers each day now makes their use possible in a policy tool, as explained in this paper.

A computer error affecting the compilation of the daily figures for new cases for England was discovered on 2 October 2020, and this resulted in undercounting new infections by 15,841 between 25 September 2020 and 2 October 2020. However, this error has now been corrected. The methods discussed in Appendix B allow the readings over that interval still to be used for the measurement of SDI.

A method has been developed that makes use of the daily new-case figures to extract a measurement of the basic reproduction number. The technique is based on a predictor–corrector method that incorporates epidemic models to allow a measurement to be made of the SDI in near-real time. The degree to which the SDI needs to be adjusted to bring about the desired outcome may then be judged, allowing policymakers to concentrate on ways of achieving that goal.

3. The input to the measuring system: Cases by Date Reported

The UK Coronavirus Dashboard[42] presents data each day on the number of new infections on the, in two forms:

(i) Cases by Specimen Date;

(ii) Cases by Date Reported.

The figure for "Cases by Specimen Date" gives the number of tests for an active virus that turned out positive on a specified date. However, nonuniform reporting delays from the test centres across the country mean that 5 days are needed before this measurement stabilizes somewhere close to its final value. Thus, for example, if one had consulted the Dashboard on 10 October 2020, one would have seen a figure well below the number of people actually testing positive on that date. It would have been necessary to wait until 15 October before a reasonably stable figure emerged for 10 October.

The figure for "Cases by Date Reported" differs from "Cases by Specimen Date" because it includes all the positive tests notified to the central information point that day, whether or not the positive results happened on that date or a few days earlier. So, for example, the figure for Cases by Date Reported on 10 October 2020, would have included some positive tests from 10 October, some from 9 October, some from 8 October and so on. Thus the measurement contains information directly relevant to the day in question plus, in addition, what amounts to extrapolated information from the previous few days. The statistic does not change and is usable on the day it is read. It gives the earliest indication of how many people are being infected and the government uses it as the headline figure to be promulgated to the media and the general public.

Strictly, the figure for Cases by Date Reported lags on that for Cases by Specimen Date, as shown in Figure 6, drawn retrospectively, after the Cases by Specimen Date had stabilized. An estimate of the time lag can be made by advancing the Cases by Date Reported in time (viz., treating them as if they had been reported earlier) by 0, 1, 2, 3 and 4 days and comparing the root mean squared error, as shown in Figure 7. The optimal lag is 2 days, which means that the current figure for Cases by Date Reported provides a good estimate of the Cases by Specimen Date two days ago. This is, of course, three days before the five days it takes for the measurement of Cases by Specimen Date to stabilize. Thus Cases by Date Reported may be regarded for practical purposes as a leading, not a lagging, indicator of Cases by Specimen Date. It is for this reason that Cases by Date Reported has been chosen as the measured parameter for feeding into the filtering system used in this paper.

4. Accounting for the widely varying numbers of viral tests

The use of the raw figure for Cases by Date Reported over medium to long time periods is rendered problematic by the very large increase in the number of Covid-19 swab tests carried out over the months since April 2020 (Figure 4). Figures 8 and 9 provide further illustration of the size of the effect. While Figure 8 shows cases rising in October 2020 to four times what they were in April 2020, Figure 9 shows the percentage of positive tests to be much lower in October

---

[42] GOV.UK, Coronavirus (COVID-19) in the UK https://coronavirus.data.gov.uk/

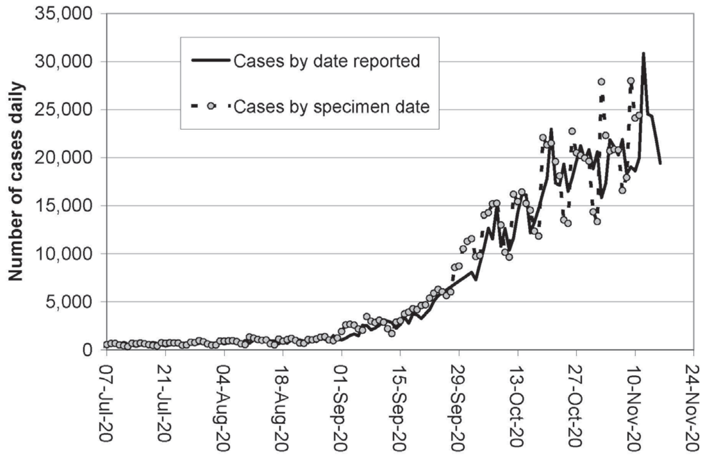

Figure 6. Cases by Date Reported, shown lagging Cases by Specimen Date.

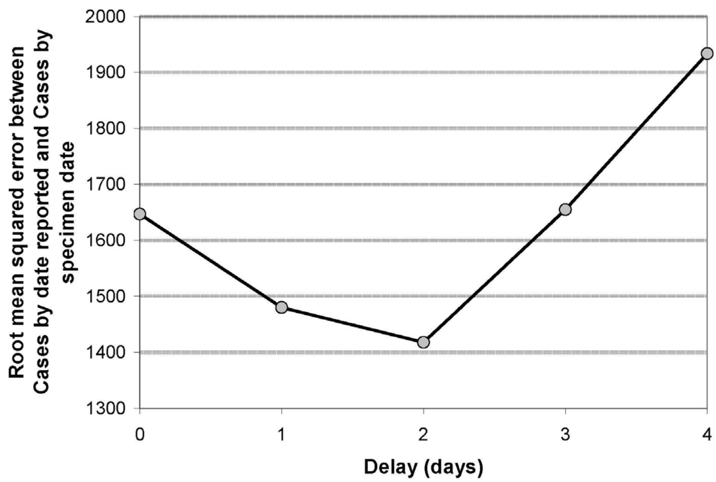

Figure 7. Comparing Cases by Date Reported and Cases by Specimen Date after allowing for a 5-day settling period for the latter. The root mean squared error between the two measurements is given as a function of the delay assumed between Cases by Specimen Date and Cases by Date Reported.

2020 than it was at the height of the first wave. The geometries of the graphs are clearly transposed; allowance needs to be made for the much smaller number of tests carried out earlier on. Appendix D provides detail on how the effect can be corrected for. The assumption is made that the reporting of tests from laboratories all over the country will be subject to delays similar to those associated with the notification of positive cases. Evidence for this comes from the UK Coronavirus Dashboard, where the number of tests is admitted to decline every weekend, but the lowest numbers recorded normally occur on a Monday, with Sunday and Tuesday being the next lowest (see Figure 10).

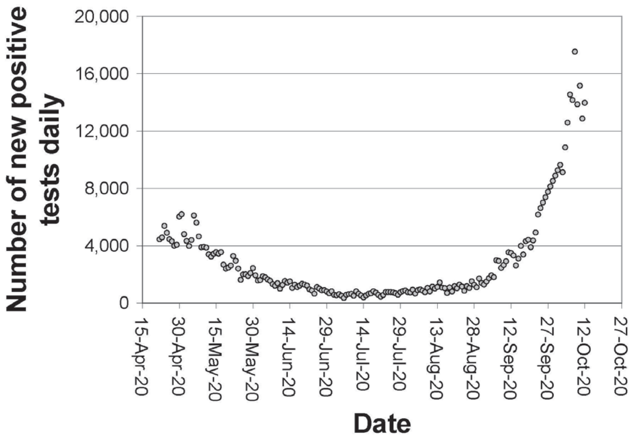

Figure 8. Number of PCR swab tests found to be positive each day.

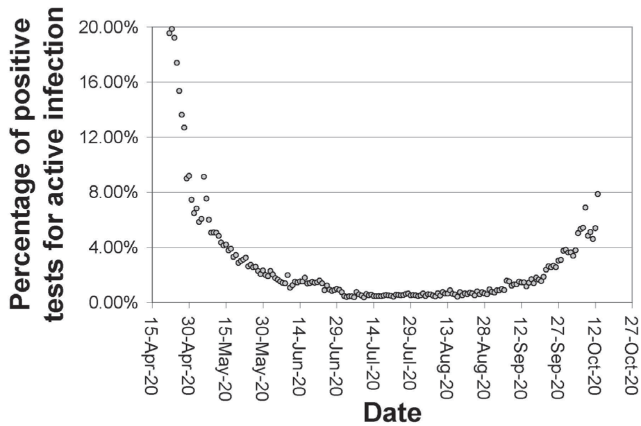

Figure 9. Percentage of PCR swab tests found to be positive.

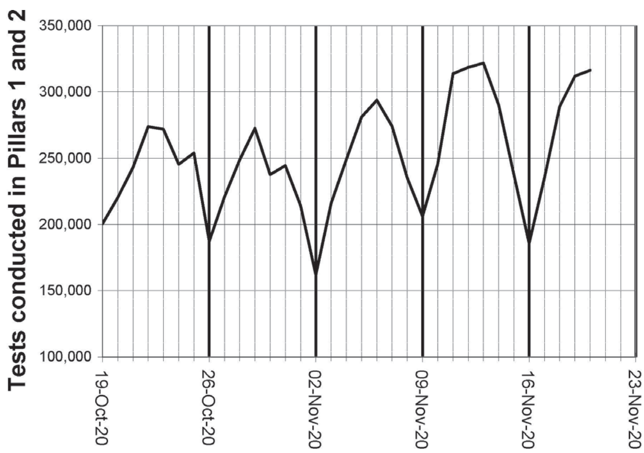

Figure 10. Periodicity of the Tests by Date Reported statistic: each low point is on a Monday.

Let the number of cases reported on date t, when $w(t)$ tests were nominally carried out, be $m(t|w(t))$, and the number of cases reported k days earlier, on date $t-k$, when $w(t-k)$ tests were administered, be $m(t-k|w(t-k))$. What would be the number $m(t-k|w(t))$ that would have been reported on the earlier date $t-k$ if the number of tests carried out on that date had not been $w(t-k)$ but the same as on the later date, namely $w(t)$? In fact, as shown in Appendix D, the number of tests may be broken down into those for Pillar 1, $w_1(t)$, and those for Pillar 2, $w_2(t)$:

$$w(t) = w_1(t) + w_2(t). \tag{1}$$

Thus the difference in the number of tests between dates $t-k$ and t is will be, for Pillar 1:

$$\Delta w_1(t, t-k) = w_1(t) - w_1(t-k). \tag{2}$$

The expected difference in cases arising from Pillar 1 may then be found from the probability on that date of obtaining a positive result from a test in that pillar, $p_1(t-k)$

$$E(\Delta M_1(t-k, t)) = p_1(t-k)\Delta w_1(t-k, t). \tag{3}$$

The estimated number of cases, $\hat{m}_1(t-k|n_1(t))$, under Pillar 1 if $w_1(t)$ tests had been carried out is then the sum of the observed value, $m_1(t-k|w(t-k))$, and the expected value from equation (3):

$$\hat{m}_1(t-k|w_1(t)) = m_1(t-k|w(t-k)) + p_1(t-k)(w_1(t) - w_1(t-k)). \tag{4}$$

In a strictly analogous way, the estimated number of cases, $\hat{m}_2(t-k|w_2(t))$, under Pillar 2 is given by

$$\hat{m}_2(t-k|w_2(t)) = m_2(t-k|w(t-k)) + p_2(t-k)(w_2(t) - w_2(t-k)). \tag{5}$$

Applying equations (4) and (5) with $k > 0$ allows a fair comparison between the number of cases that would have been expected on an earlier date and the number at a later reference date. Using the same equations with $k < 0$ allows the number of cases that would have been expected on a later date to be compared fairly with the number on an earlier reference date. The application of this method to the Cases by Date Reported for England from 31 March 2020 to 20 November 2020 produces the graph shown in Figure 11. The apparent relative magnitudes of the first and the second wave are reversed.

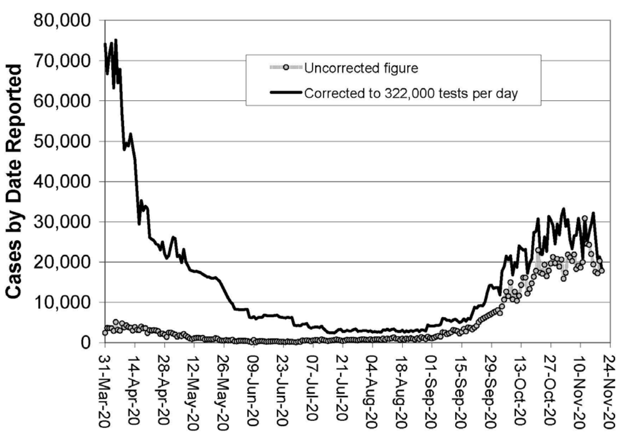

Figure 11. Correcting Cases by Date Reported for the number of tests carried out.

The method also provides a better understanding of the magnitudes of daily incidence figures that are close to one another in time. Thus the uncorrected time series shows the infections for 12 November 2020 as towering above those either side of it. But correcting for some of the very significant periodicity in the number of tests recorded suggests that there would be other comparable numbers of daily new cases just before and just afterwards (see Figure 12). These ideas receive quantitative support from the maxima of the 7-day averages for corrected cases and deaths in England during the first wave in March–April 2020 and the second wave in October–November 2020 (see Table 1).

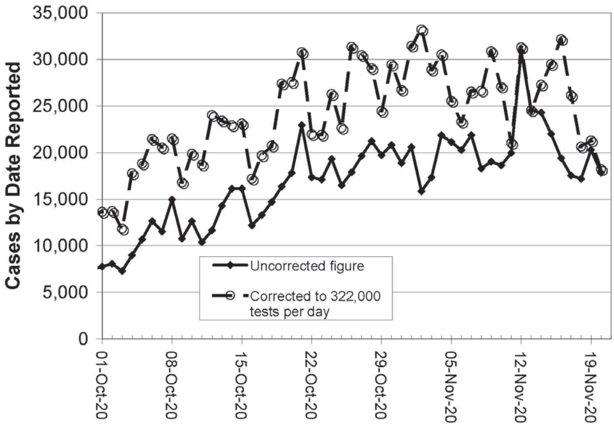

Figure 12. Detail from Figure 11, showing how the apparent outlier on 12 November becomes just one of a number of high peaks when allowance is made for the number of tests carried out each day.

Table 1. Peak 7-day averages for deaths and cases corrected for the scale of testing during the first wave (April 2020) and the second wave (November 2020).

	End date	Peak 7-day average in wave	Ratio of peaks
Cases	06-Apr-20	69,959	
	05-Nov-20	30,001	0.43
Deaths	10-Apr-20	876	
	21-Nov-20	392	0.45

The ratio of the peak 7-day average of corrected daily cases during the second wave to the equivalent number in the first wave is 0.43. Meanwhile the corresponding ratio for deaths is almost identical at 0.45. An alternative way of viewing the data is to take the ratio of the peak 7-day average of deaths to the peak 7-day incidence rate as measured by the corrected Cases by Date Recorded. This produces approximate figures for nominal infection fatality rates during the two waves: 1.25% in the first wave, 1.3% in the second. In fact, as will be discussed in §6.2, there are likely to be two unnotified cases for every one recorded, so that the true figures for infection fatality are likely to be roughly 0.4%, providing confirmation for the lower end of the range quoted in Section 2.1. The analysis suggests that, unfortunately, the fatality rate for Covid-19 does not appear to have improved in the 7 month interval since April 2020.

5. The effect of T-cells on immunity

Scientists at Sweden's Karolinska Institute reported in August 2020 their finding that twice as many people acquired immunity from Covid-19 through the action of their T-cells than via antibodies (which are made by B-cells). Their conclusion was based on a study of those who had recovered from the disease in the first phase of the coronavirus pandemic (roughly April and May 2020 in Sweden) together with those donating blood in May 2020.[43,44] The potential importance of the Karolinska result led the author of the present paper to call for UK laboratories to confirm the T-cell effect as a matter of urgency.[45] The results from just such a study have now been reported:[46,47] analysis by Public Health England and university partners of a population of 2,197 seronegative UK key workers (from the police, fire service and health services) starting in June 2020 showed that:

(i) high levels of SARS-CoV-2-responsive T-cells are associated with protection from symptomatic SARS-CoV-2 infection; and

(ii) 12.9% (95% confidence interval: 11.5% to 14.3%; see Appendix F) of the population had such T-cell levels and had protection despite being seronegative, that is to say almost certainly not having contracted Covid-19. It was most likely that they harboured SARS-CoV-2 responsive T-cells that had been primed by common-cold coronaviruses rather than SARS-CoV-2.

These results from Public Health England are consistent with and confirm the Karolinska Institute findings. Sweden's Public Health Authority found the fraction of Sweden's population with Covid-19 antibodies was 6.3% in the last week of May 2020[48] in the nine counties of Jämtland, Jönköping, Kalmar, Skåne, Stockholm, Uppsala, Västerbotten, Västra Götaland and Örebro, with Stockholm registering 7.3%.[49] Doubling these figures in line with the Karolinska Institute finding gives a range 12.6–14.6%, which brackets the Public Health England figure. The new findings suggest that on 29 June 2020, when the ONS reported that 6.3% of England's

---

[43] Karolinska Institute, Immunity to COVID-19 is probably higher than tests have shown (18 August 2020) https://news.ki.se/immunity-to-covid-19-is-probably-higher-than-tests-have-shown

[44] Lourenço, J., Pinotti, F., Thompson, C. and Gupta, S., The impact of host resistance on cumulative mortality and the threshold of herd immunity for SARS-CoV-2 (16 July 2020) https://doi.org/10.1101/2020.07.15.20154294

[45] Thomas, P., Driving the economy into recession is killing more people than Covid ever could. *Mail Online* (29 August 2020) https://www.dailymail.co.uk/debate/article-8677207/Driving-economy-recession-killing-people-Covid-could.html

[46] Wyllie, D. et al., SARS-CoV-2 responsive T cell numbers are associated with protection from COVID-19: A prospective cohort study in keyworkers. medRxiv preprint (2020) https://doi.org/10.1101/2020.11.02.20222778

[47] Public Health England, 2020, Evaluating detection of SARS-CoV-2 antibodies. EDSAB-HOME study: details, research protocol and outputs, https://www.gov.uk/government/publications/evaluating-detection-of-sars-cov-2-antibodies-at-home-study

[48] Sweden COVID-19 studies: Ongoing infection, antibody tests in blood donors (19 June 2020) http://outbreaknewstoday.com/sweden-covid-19-studies-ongoing-infection-antibody-tests-in-blood-donors-98648/

[49] Pandemic Preparedness, About 7% of people have antibodies to Covid-19 (22 May 2020) https://pandemic.internationalsos.com/reports/about-7-of-people-in-stockholm-have-antibodies-to-covid-19-may-22-2020?page=3

population had antibodies[50] and therefore immunity, the true figure for immunity was 6.3% + 12.9% =19.2%. The figure for London would have been higher at the beginning of July 2020, perhaps in the range 25% to 30%, which would account for the capital being less badly affected by the second wave than areas in the North of England, which had not experienced so many cases in the first wave.

6. The predictor–corrector measurement system

6.1 Predictor

Using the single cohort model described in Appendix A of a previous paper,[26] the first equality in equation (A.13) gives the rate of change in the number n of active cases per day, dn/dt, which may be written:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{\mathrm{d}n_x}{\mathrm{d}t} - \frac{\mathrm{d}n_r}{\mathrm{d}t} \tag{6}$$

where $\mathrm{d}n_x/\mathrm{d}t$ is the number of new cases per day while $\mathrm{d}n_r/\mathrm{d}t = n/\tau_{\mathrm{inf}}$ (equation A.9 of ref. 26) is the rate of change in the number of people who have recovered or, in a minority of cases, died. The term $\mathrm{d}n_x/\mathrm{d}t$ is set to the figure for Cases by Date Reported, after correction for the number of tests per day. Numerical integration over a period of one day gives:

$$n(t) - n(t - \delta t) \approx \frac{\mathrm{d}n_x}{\mathrm{d}t}\delta t - \frac{n}{\tau_{\mathrm{inf}}}\delta t \tag{7}$$

where τ_{inf} is the average time between someone becoming infected and passing on the infection to others. This is taken to be 8.3 days, the weighted average of the corresponding values in the two-cohort model used in the Base Case by Thomas (2020).[36] With $\delta t = 1$ day, equation (7) may be re-expressed to give the number of active cases on date t_n as:

$$n(t_n) \approx n(t_{n-1}) + \frac{\mathrm{d}n_x}{\mathrm{d}t} - \frac{n}{\tau_{\mathrm{inf}}}. \tag{8}$$

The predictor accounts only for reported cases, not those unreported. It is appropriate to use the figure for $n(t_n)$ at the requisite t_n as the initial condition for the number of active cases n_1 in Cohort 1 in the corrector model, as Cohort 1 is concerned only with notified infections.

Using equation (A.14) from Appendix A of ref. 26, we may also express the rate of change per day of the number of people becoming infected as

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{n}{\tau_{\mathrm{inf}}}\left(R_0 \frac{n_s}{N} - 1\right) \tag{9}$$

where N is the total population, n_s is the number of susceptible individuals in the population and R_0 is the *basic* reproduction number (or social distancing index).

[50] ONS, Coronavirus (COVID-19) Infection Survey pilot: England (17 July 2020) https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/england17july2020#antibody-data

Clearly n_s will decrease over time as more individuals are infected and recover, so we may define an *effective* reproduction number or "R-rate", $R(t)$, as

$$R(t) = R_0 \frac{n_s(t)}{N}. \tag{10}$$

Substituting from equation (10) into equation (9) gives:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{n}{\tau_{\mathrm{inf}}}\left(R(t) - 1\right) \tag{11}$$

or

$$\frac{\mathrm{d}n}{n} = \frac{1}{\tau_{\mathrm{inf}}}\left(R(t) - 1\right)\mathrm{d}t. \tag{12}$$

Equation (12) may be integrated analytically from a beginning time t_b to a later time t by assuming that $R(t)$ stays approximately constant over the integration interval $t - t_b$. This then produces:

$$\ln \frac{n}{n(t_b)} \approx \frac{1}{\tau_{\mathrm{inf}}}(R - 1)(t - t_b). \tag{13}$$

Thus the R-rate estimated by the predictor at each date is

$$R(t)\big|_{\mathrm{pred}} \approx 1 + \frac{\tau_{\mathrm{inf}}}{t - t_b} \ln \frac{n(t)}{n(t_b)} \tag{14}$$

where $n(t)$ is the number of people with an active Covid-19 infection at date, t, and $n(t_b)$ is the corresponding number a few days before, at date, t_b. Given that new case numbers arrive daily, the minimum possible value for $t - t_b$ would be 1.0. But such a setting would be highly susceptible to the noise apparent from examining the raw measurements. Setting the interval as 7 days will make automatic allowance for the marked weekly periodicity seen in Figure 10 and reduce measurement noise more generally:

$$t - t_b = 7. \tag{15}$$

The values generated for $R(t)|_{\mathrm{pred}}$ over the range of dates of interest may then be transferred to the corrector.

6.2 Corrector

Here the two-cohort model[51] of those infected is used, where Cohort 1 contains those who experience recognizable symptoms and are tested and found positive. Meanwhile the members of Cohort 2 experience less severe symptoms or are asymptomatic and are not tested, even when they become infected. The parameters characterizing the two groups may be found in Table 1 of Thomas (2020)[36] and are as follows. For Cohort 1, the basic reproduction number at the start of

[51] Thomas, P., The length and severity of the coronavirus recession estimated from the dynamics of relaxing lockdown. *Nanotechnology Perceptions* **16** (2020) 100–129 (http://www.colbas.org/ntp/opnAxs/N07TH20A.pdf).

the epidemic is taken to be $R_{01}(0) = 2.53$, while the average time between infection and transmission is given by $\tau_{\mathrm{inf},1} = 8.82$ days. The corresponding figures for Cohort 2 are $R_{02}(0) = 1.87$ and $\tau_{\mathrm{inf},1} = 8.25$ days.

The original model allowed for a fraction, $\theta_1 = 11.6\%$, of the total number of cases being reported, with $\theta_2 = 1 - \theta_1 = 88.4\%$ left unreported. These percentages, which are based on an optimal match to daily new cases figures up to 30 April 2020,[36] are similar to figures presented by Petersen and Phillips[52] for the fraction of cases with recognized Covid-19 symptoms between 27 April 2020 and 27 June 2020. They found that 86.1% of a random sample who tested positive reported no symptom specific to Covid-19 while 76.5% reported no symptoms at all.

Two new factors mean that the fraction θ_1 of people assigned to Cohort 1 should be updated. First there is the enormous rise in testing in England between April 2020 and November 2020, which increased the fraction of cases with symptoms that are known about (these correspond to Cohort 1). Figure 13 compares the Cases by Date Reported for England published on the UK's Coronavirus Dashboard with the figure for new cases per day estimated for England by the ONS Coronavirus (COVID-19) Infection Survey pilot. The slope of the least-squares fitting line in Figure 13 suggests that 34% of all new infections were being discovered between July and November 2020.

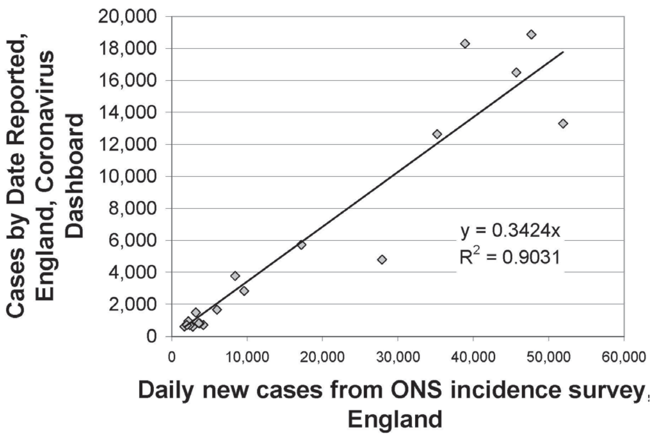

Figure 13. Comparison of new cases reported on Coronavirus Dashboard against the estimate of new cased from ONS surveys, July to October 2020.

The second factor is the recent knowledge that 12.9% of the population has prior T-cell immunity. This means that a large number of English people, viz. $0.129N$, where N is the 56.5 million-strong population of England, will never develop the virus antigens necessary for the ONS to identify an active infection. We may deduce the infected people identified by the ONS survey would need to have been drawn from a population that comprised only a subset

[52] Petersen, I. and Phillips, A., Three quarters of people with SARS-CoV-2 infection are asymptomatic: analysis of English household survey data. *Clinical Epidemiology* (9 October 2020) file:///C:/Users/Philip/Downloads/clep-276825-three-quarters-of-people-with-sars-cov-2-infection-are-asymp.pdf

$N' = 0.871N$ of the population at the outset of the epidemic. Only a fraction, identified as 0.3424 from Figure 13, of the active new cases in the population will be tested under the Government scheme and found positive, even though there are a further 65.76% of the population who have Covid-19 and would be found positive if tested for antigens. This is probably because these people are asymptomatic or are only mildly symptomatic, despite having an active infection. Whatever the reason, roughly two thirds of the population without T-cell immunity will not come forward for test.

The number of people who would come forward and be tested after contracting the virus may be estimated as $0.3424N' = 0.3424 \times 0.871N = 0.3N$. These are the people in Cohort 1. Thus the revised value for the fraction θ_1 is 30%. The remaining fraction of the population will be assigned to Cohort 2, so that $\theta_2 = 1 - \theta_1 = 70\%$. Those who are T-cell immune can be added to the number of people in Cohort 2 who have recovered from Covid-19. The value of the overall R-rate in the corrector module is found by multiplying the predicted value, $R(t)|_{\text{pred}}$, by a modification factor, f_R (adjusted to give a best fit to the data on Cases by Date Reported). Multiplying the result by $N/(n_{s1}(t) + n_{s2}(t))$, where n_{s1} is the number of susceptible individuals in Cohort i at date, t, gives the overall SDI, $R_0(t)$:

$$R_0(t) = f_R \frac{N}{n_{s1} + n_{s2}} R(t)\Big|_{\text{pred}} . \tag{16}$$

It is assumed that each cohort conforms to the social distancing restrictions in a similar way, leading to proportionate reductions in the SDIs R_{01} and R_{02}. Thus the ratio r_R of the two SDIs will stay the same as restrictions change with time:

$$r_R = \frac{R_{02}(t)}{R_{01}(t)} = \frac{R_{02}(0)}{R_{01}(0)} = \text{constant} \tag{17}$$

where $r_R = 1.87/2.53 = 0.74$. A combined SDI, $R_0(t)$, may be found as the weighted average:

$$R_0(t) = \theta_1 R_{01}(t) + \theta_2 R_{02}(t) = \theta_1 R_{01}(t) + r_R \theta_2 R_{01}(t) \tag{18}$$

where equation (17) has been used in the second line. The individual SDI for Cohort 1 may therefore be found from the overall number by:

$$R_{01}(t) = \frac{R_0(t)}{\theta_1 + r_R \theta_2} \tag{19}$$

allowing $R_{02}(t)$ to be then calculated using equation (17). The two-cohort model requires initial values for its state variables, $n_1(t_0)$, $n_2(t_0)$, $n_{r1}(t_0)$ and $n_{r2}(t_0)$, on the chosen starting date, $t_0 = 4$ July 2020.

As discussed in the section on the predictor, $n_1(t_0)$ may be taken as equal to the predictor's calculated value for $n(t_n)$ given by equation (8), when $t_n = t_0$. The ONS survey suggested that 6.3% of the English population (3.56 million people) had antibodies against Covid-19 on 29 June 2020, and it is assumed that the same number applies on the starting date, $t_0 = 4$ July 2020. These people who have recovered will be split evenly between Cohorts 1 and 2. Hence $n_{r1}(t_0) = \theta_1 \times 3.56 \times 10^6 = 1.07 \times 10^6$. Meanwhile the number of people in the recovered category in Cohort 2 will be the number of people who have had Covid-19 and recovered, $(1 - \theta_1) \times 3.56 \times 10^6 = 2.49 \times 10^6$ plus the number of people who have T-cell immunity, $0.129 \times 56.5 \times 10^6 = 7.29 \times 10^6$.

Hence $n_{r2}(t_0) = 2.49 \times 10^6 + 7.29 \times 10^6 = 9.78 \times 10^6$. Finally, $n_2(t_0)$ is taken to be a multiplier r_n times $n_1(t_0)$; viz., $n_2(t_0) = r_n n_1(t_0)$.

The model is then optimized by adjusting both the multiplier r_n and the R-rate modification factor f_R so as to minimize the sum of the squared errors between model-calculated numbers of new cases, $\mathrm{d}n_x(t)/\mathrm{d}t$, and the corrected measurements, $\mathrm{d}\tilde{n}_x(t)/\mathrm{d}t$, over the period from 4 July 2020 to 26 November 2020.

7. Validation of the predictor–corrector system against independent empirical data

Optimizing the predictor–corrector system produces a value for the R-rate modifier of $f_R = 0.97$, while the optimal ratio r_n of the number of active infections in Cohort 2 to those in Cohort 1 was $r_n = 3.5$. The former is obviously close to unity, showing good agreement in the estimation of $R(t)$ between the predictor and the corrector models used. The latter is somewhat above the ratio $\theta_2/\theta_1 = 1.92$, which is consistent with the ratio of undiscovered to reported cases being substantially higher early on in the epidemic, when there was considerably less testing.

Figure 14 shows a good fit between the model to the data on daily cases presented as Cases by Date Reported over the roughly 4-month period from the beginning of July to the end of November 2020.

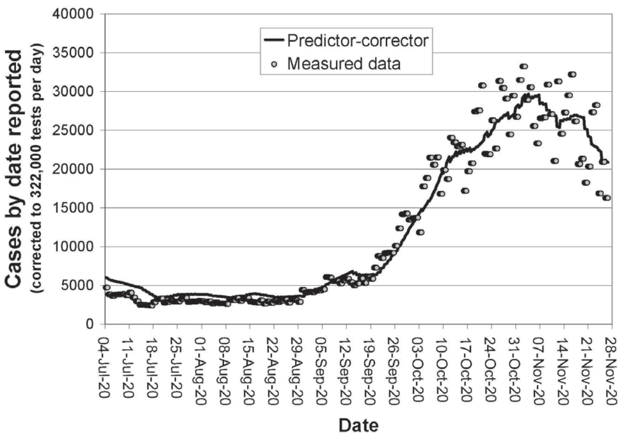

Figure 14. The two-cohort model matched to the Cases by Date Recorded after correction to the number of tests, 322,000, conducted on 13 November 2020.

A check on the validity of the predictor–corrector system is offered by an independent measurement of the number of people with an active infection at different times. The Office for National Statistics publishes weekly the results of an ongoing survey to find how many people are suffering an active infection, with about 120,000 PCR swab tests administered every fortnight.[53]

[53] ONS, Coronavirus (COVID-19) Infection Survey pilot: England, Wales and Northern Ireland (9 October 2020) https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/englandwalesandnorthernireland9october2020#covid-19-infection-survey-data . Earlier reports may be found via this URL.

Estimates are given of the number of people in England with an active coronavirus infection between 8 and 12 days before, together with 95% confidence intervals. Comparing the predictor–corrector output and the ONS survey spot measurements of the number of people in England with an active Covid-19 infection provide a diverse test for the predictor–corrector model against independently gathered empirical data (see Figure 15).

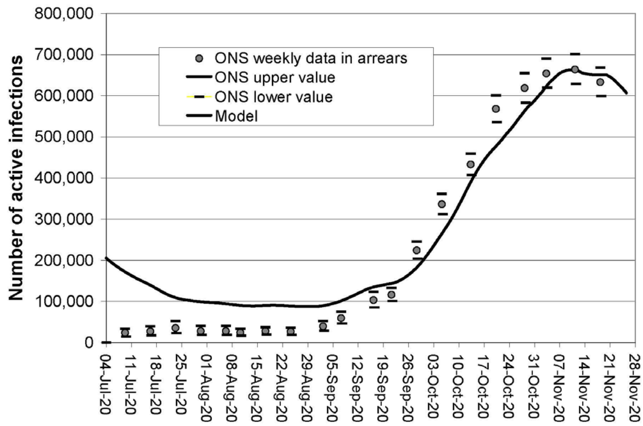

Figure 15. Comparison between the predictor–corrector output and the ONS survey measurement of the number of people in England with an active Covid-19 infection over time.

As noted in §3, the figure for Cases by Date Reported to which the model is matched lags on the specimen date by about 2 days. In addition, there will have been a delay between the date of infection and the specimen date, since symptoms typically begin in symptomatic cases between 2 and 14 days after exposure to the virus,[54] but possibly even as early as 1 day afterwards.[55]

A way of accounting for the total delay is to assume that the dates associated with the ONS survey data are correct and to estimate the delay between the ONS figure for active infections and that produced by the predictor–corrector system. Figure 16 shows the mean squared error between the two measurement methods over the period from 1 August to mid-November 2020, when the delay is varied between 1 and 7 days. There is a well defined minimum when the delay is 4 days. Bringing the predictor–corrector measurement 4 days earlier in time allows a direct comparison of the resulting profile of cases against time against ONS estimates for given reference dates (see Figure 17).

Excellent agreement is evident between the predictor–corrector output and the ONS data from the beginning of October 2020 onwards, while it is clear that the two measurement methods are capturing the same trends in the previous months. The predictor–corrector model

[54] Centre for Disease Control and Prevention, Symptoms of Coronavirus (13 May 2020) https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fabout%2Fsymptoms.html

[55] SAGE, SPI-M-O: Consensus statement on 2019 novel coronavirus (2019-nCoV) (3 February 2020) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/882710/15-spi-m-o-consensus-statement-03022020.pdf

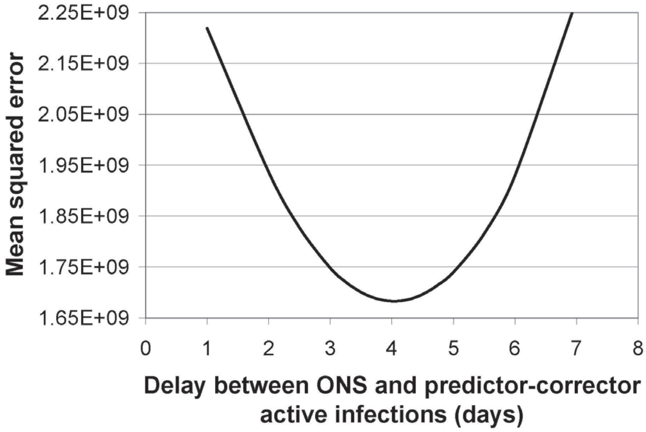

Figure 16. The mean square error as the delay is varied between the ONS and predictor–corrector figures for active infections.

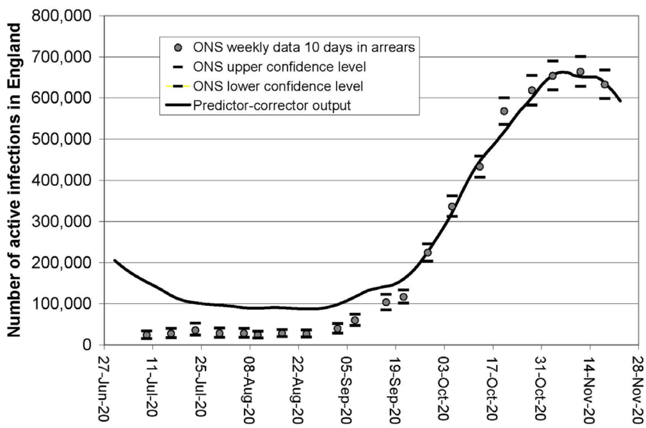

Figure 17. Comparing the ONS survey measurement of the number of people in England with an active Covid-19 infection over time with the predictor–corrector output brought forward in time by 4 days.

may thus be regarded as having passed an important validation test. The close alignment of the dynamic behaviour of the predictor–corrector model with that observed in the ONS survey data provides further confirmation for the fraction of people in the population with prior T-cell immunity, as discussed in §5.

An explanation for the disparity in the absolute number of active infections suggested by the two measurements in July and August 2020 when cases were flatlining may lie in the representation of a distributed system by a point model. Both the predictor–corrector and the ONS survey measurements seek to quantify the number of people with an active infection across the whole of England. This lumped-parameter approach is more likely to be accurate when there are more cases, which will increase the probability of large parts of the country being affected and thus give greater geographical uniformity of infections.

The ONS survey figures for active cases are, of course, retrospective and already 10 days in arrears on the day of publication. The decision-maker needs to wait 7 days between sets of figures, so that the average delay for any given date is about 14 days. It has been shown in this Section that the predictor–corrector method is able to produce figures of comparable accuracy on a daily basis with a delay of only 4 days, a gain of ten days, on average on the waiting time associated with the production of an estimate based on the ONS survey.

8. Deducing the R-rate using data from ONS weekly Covid-19 infection surveys

The ONS data may also be used to generate an independent figure for the R-rate. The survey values for active infections are set between 4 and 9 days apart, allowing equations (14) and (15) to be applied to calculate a value for the R-rate, with $t - t_b$ set to the length of the relevant time interval between measurements. Because these are deduced from survey sampling, each needs to be treated as a random variable from which an expected value and a 95% confidence interval may be found. The procedure is outlined in Appendix C.

9. Comparison of the R-rates produced by the predictor–corrector with those based on the ONS survey

Figure 18 compares the ONS survey values against the nominal R-rate measured by the predictor–corrector method, where "nominal" is taken to mean on the date of measurement. There is good general agreement between the two measurements overall, with very close agreement from the beginning of October 2020.

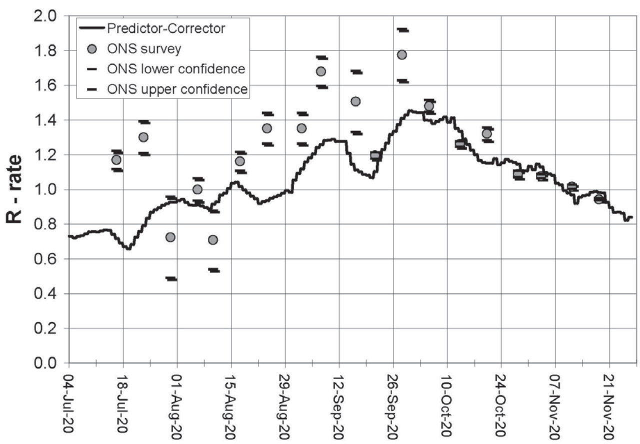

Figure 18. Comparison of the R-rate measured by the predictor–corrector method and by the ONS survey method.

Figure 19 shows the situation where the R-rate calculated by the predictor–corrector is shifted earlier in time by 4 days to give better comparability between dates. The match is again very good from early October 2020 onwards.

The ONS-based R-rates are significantly more erratic from July 2020 to the end of September 2020 than those produced by the predictor–corrector model. The error bars are, of course, very

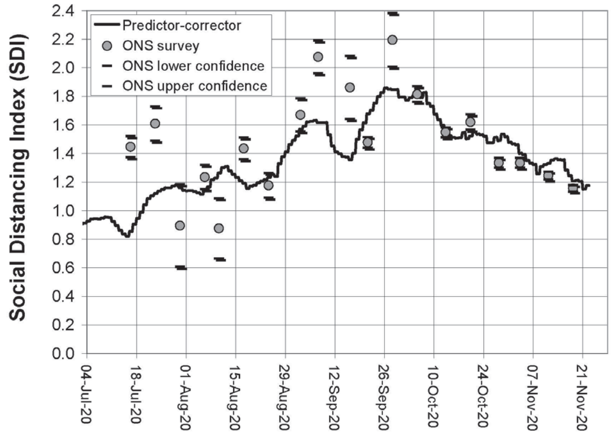

Figure 19. Comparison of ONS survey results with the predictor–corrector measurements shifted earlier in time by 4 days.

much larger in this region, but full use of them would be needed to restrain the variation to a level similar to that suggested by the predictor–corrector model. The relatively large variations might be flagging the difficulty of choosing a fully representative sample when cases are very low: more infections might be ongoing in some parts of the country than in others. Whether or not this is the explanation, the match improves significantly when the number of active cases is larger.

Figure 20 gives a detail of Figure 19 covering the period from the beginning of October 2020 to the end of November 2020. The predictor–corrector suggests that England's R-rate was falling more or less monotonically from early October 2020 onwards and had already dropped below 1.0 on 5 November 2020, the first day of England's second lockdown.

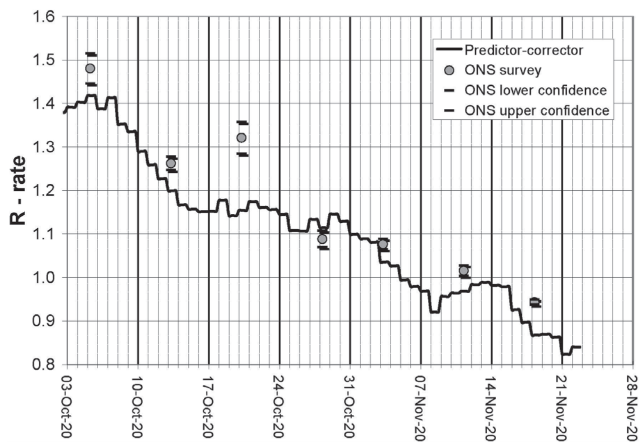

Figure 20. Comparison of ONS survey results with the predictor–corrector measurements shifted earlier in time by 4 days: detail.

The ONS survey measurements suggest that the R-value was about 1.08 on 2 November 2020, with $R = 1$ lying within the 95% confidence interval on 11 or 12 November 2020, about a week after the second English lockdown began.

Both ONS-based and predictor–corrector measurements show a strong falling trend throughout October 2020, which could mean that further restrictions were not needed in order to satisfy the government's goal of reducing the R-value below 1.0. Such a conclusion was reported by Spector, based on results of the ZOE app, which tracks the symptoms reported by a million users on a weekly basis.[56]

10. Using ONS survey data to measure the social distancing index, R_0

Dividing the R-rate by the fraction of people in the population who are still susceptible, in line with equation (10), eliminates the extraneous and confounding extra factor and allows proper concentration on how well people are modifying their behaviour to reduce the possibilities for infection. The result is, of course, the SDI, R_0.

The ONS collects survey data not only on the number of people with an active infection, as referred to in §7, but also on the fraction of people in the population with antibodies. To this needs to be added the measured fraction of the population with T-cell immunity to infection, as discussed in §5, so as to find the total level of immunity in the population. The fraction who are still susceptible to Covid-19 is then the complement of the fraction with immunity.

As the percentages of people with antibody immunity and T-cell immunity have both been found from surveys, each of these fractions needs to be regarded as a random number. The randomness carries over to the R-rate found from these survey figures as it, too, is derived ultimately from survey data. 95% confidence levels are again appropriate, and a derivation of the necessary method for their calculation is given in Appendix E.

11. Comparison of the values produced by the predictor–corrector and from the ONS surveys for the social distancing index

Figure 21 shows the predictor–corrector output for the SDI advanced 4 days earlier to enable a direct comparison with spot values computed from ONS survey figures. Once again there is good general agreement, especially when the error bars on the ONS-based estimates are shorter, from the start of October 2020 onwards.

The ONS-based data points display a high degree of noisiness before that epoch, which might not be fully representative of the infection status of the 56.5 million people making up the population of England, in the sense of not reflecting the inertia that such a large body of people would likely exhibit. Such inertia is, of course, already implicit in the model used to take the predictor–corrector measurement. The more erratic behaviour of the ONS data points before October may well be connected with the number of people sampled being initially low but then increasing over time. The initial batch of invitations to take part in the infection survey in May

[56] Spector, T., The second wave appears to have peaked before lockdown. *The Spectator* (29 November 2020) https://www.spectator.co.uk/article/the-second-wave-appears-to-have-peaked-before-lockdown#

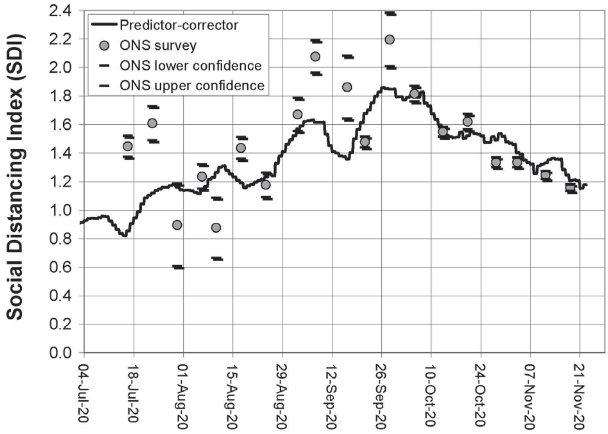

Figure 21. Measured value of the social distancing index, R_0: comparing ONS figures, after allowing for T-cell immunity, with the predictor–corrector measurement advanced by 4 days in time.

2020 led to the eventual enrolment of 10,326 households. This number had risen to 32,676 by 21 August 2020 and to 153,500 households by 8 October. The corresponding pool of people eligible to be sampled rose from 69,391 on 21 August 2020 to 341,534 by 8 October 2020.

Based on the predictor–corrector measurement, England's SDI rose above 1.0 in the middle of July 2020, then held fairly steady around 1.2 for most of August 2020, before rising rapidly to 1.6 in the last week of that month. The SDI fell back to about 1.4 by the middle of September 2020, but then rose again to reach 1.8 in the third week of that month. It stayed there for a fortnight before starting a long fall on about 7 October 2020.

The period from the middle of September to early October 2020 coincided with the reopening of universities to new students, followed first by freshers' week and then the return of existing students. Six months into the pandemic, university students may well have been aware of the low level of risk they faced from contracting Covid-19 as a result of their youth, as discussed in Section 1.1. There may well have been enhanced mixing at the beginning of the new academic term, possibly less than normal, but nevertheless more than the government would have wished. This could have been a substantial contributor to the significant but temporary increase in the SDI.

The SDI can be seen to have maintained a fairly steady level around 1.5 in the week immediately before and after the Tier restrictions were imposed on 17 October. However, the R-rate continued to fall during this time as a result of the ongoing reduction in the number of people still susceptible. The fall in the SDI recommenced at the end of October 2020, so that the SDI was about 1.3 when the second English lockdown was imposed on 5 November 2020. As discussed in §9, the R-rate may well have been already below 1.0 at that time. The SDI maintained its fall through November 2020, reaching 1.18 by 22 November 2020.

12. Scenarios to the end of the epidemic in England

The principal purpose of the predictor–corrector system is to provide accurate measurements of the SDI on a daily basis, together with the R-rate and the number of active infections. The SDI gives both people and government the earliest possible measurement of how well people are keeping to the rules, and is therefore suitable for control purposes.

However, the model used in the measurement system can be used to give a "look-ahead" facility. The two-cohort model enables forward predictions to be made of the number of daily cases likely to be reported for a specified profile of the SDI over future time.

The government's planned mass vaccination campaign is expected to start with NHS staff in the week beginning 7 December 2020.[57] For this forward projection, it is assumed that full-scale inoculation of the public will begin 3 weeks later on 28 December 2020 and continue, at the rate of 5 million per month, until 95% of the population has completed the course of injections (the first inoculation is followed by a booster). It is assumed that priority will be given to the most vulnerable in the manner described in Appendix G, that the vaccine will be 70% effective, and that it will be 28 days after the first dose before immunity is conferred. Appendix F describes the additional modelling needed to account for vaccination, with Appendix G allowing for the effects of vaccines being given preferentially to the most vulnerable first.

§7 showed that the predictor–corrector measurements of the number of active cases, the SDI and the R-rate are all strictly subject to a 4-day delay, so that the measurement taken today refers, strictly, to the situation four days ago. However, the timescales for the scenarios considered are sufficiently long to render negligible the error caused by using nominal rather than true dates. Hence nominal dates will be quoted.

12.1 Scenario 1. The R-rate is kept constant at *R* = 0.95 (except for 5 days at Christmas) and restrictions are eased continuously, implying a steadily increasing SDI, until all are removed

This case is an idealization of the government's policy of keeping the R-rate below 1.0, except for the 5 days at Christmas. The set point for the R-rate was chosen to be 0.95, just below 1.0, to lessen the harmful impact of the restrictions on society, on the economy, and on the nation's health, while still conforming to the government's frequently declared wish for an R-rate below 1.0 at all times. A lower value would reduce the incidence of Covid-19 but at the expense of increased collateral damage.

The R-rate is assumed to take the value, 0.84, recorded for 26 November 2020 for the remainder of the lockdown period and then rise to 0.95 on 2 December 2020. The SDI is then found from equation (10) by treating the R-rate as the exogeneous variable. Since the number $n_s(t)$ of susceptible people is falling, the SDI will start to rise, despite the R-rate remaining constant (see Figure 22).

For the 5 days from 23 to 27 December 2020, when lockdown rules are to be relaxed, the SDI is assumed to take the maximum value measured since England relinquished the second lockdown, namely 1.86, found on 23 September 2020. The R-rate depends on the SDI for this short period and is calculated from equation (10). Note that the R-rate will be lower than it was in September 2020 because more people have contracted the disease and the number of susceptible people is thus smaller—down to ~70% of the population by Christmas 2020.

---

[57] Campbell, D., Hospitals in England told to prepare for Covid vaccine rollout in 10 days' time. *The Guardian* (27 November 2020) https://www.theguardian.com/world/2020/nov/27/hospitals-england-told-prepare-early-december-covid-vaccine-rollout-nhs

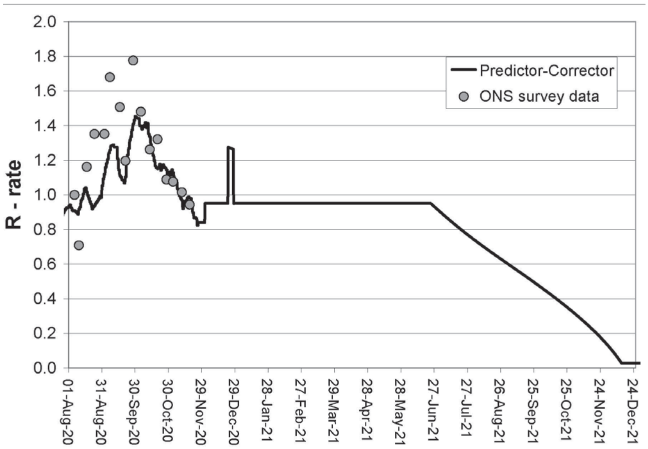

Figure 22. The R-rate for Scenario 1, §12.1.

The R-rate is assumed to return to 0.95 on 28 December 2020, when the SDI resumes its role in this scenario as the dependent variable and, following its fall immediately after Christmas, begins to climb, a rise that continues throughout the first half of 2021. It reaches its unrestricted value, conservatively estimated to be 3.0, at the end of May 2021. Once this limit is reached, the R-rate becomes the independent variable in equation (10) and begins its monotonic decline. It descends to almost zero by December 2021, at which point the epidemic is over in England.

If it turned out that some restrictions were still in place when the SDI reached 3.0, then the process of gradually easing constraints would continue until all had been removed, perhaps at an SDI of 3.5. This later time would then mark the point when the R-rate became the dependent variable. Conversely, if all restrictions were gone at some earlier time when the SDI was 2.5, say, then the R-rate would become the dependent variable before the end of May. A further possible scenario is the emergence of new, more transmissible strains that might have the effect of raising the maximum SDI to 4.5, say, if they supplanted the original strain entirely. The principles laid out here would remain the same, but full removal of restrictions would be delayed.

Figure 23 shows the behaviour of the SDI. The assumption embedded in this scenario is that the government will maintain the R-rate at 0.95, which will entail relaxing restrictions continuously from the beginning of January 2021. The vaccination effect becomes visible from the end of January onwards, reflected in the steepening of the upward slope.

Figure 24 shows the number of active infections dropping to 540,000 by the end of lockdown and then falling further to half a million by late December 2020 before there is an increase to just under 600,000 just after the Christmas mixing. The number of people with an active infection then falls continuously in the new year, dropping to 300,000 by early June 2021, when society is declared fully open.

The SDI is at its maximum value of 3.0 at this point, and cannot be increased any further to maintain the R-rate at 0.95. Hence the latter is allowed to drop naturally from this point, and the epidemic now wanes. There are just 100 active cases across England by the end of October 2021, and the epidemic is to all intents and purposes already over.

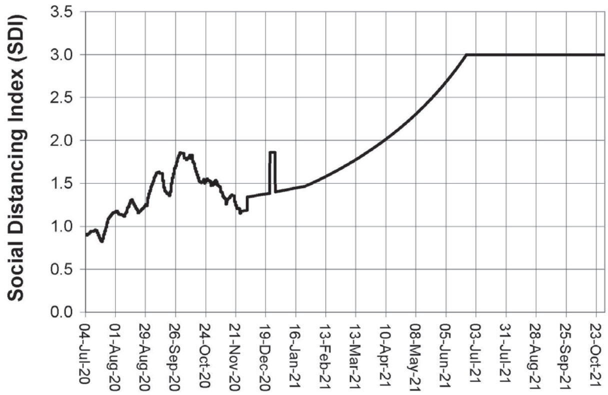

Figure 23. Social distancing index over time for Scenario 1, §11.1.

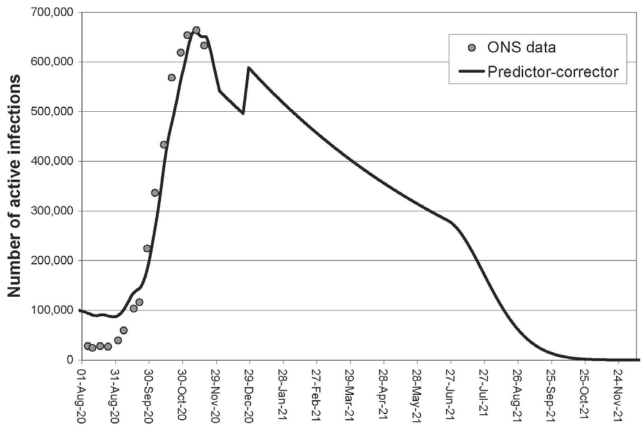

Figure 24. The number of active infections over time, Scenario 1, §11.1.

Figure 25 shows the alternative perspective of Cases by Date Reported, corrected to 322,000 tests per day. There is an uptick in daily cases upon emergence from lockdown, then a decline in December 2020 before a spike occurs as a result of the 5-day Christmas relaxation of restrictions, when cases per day briefly go back up to over 30,000. The excursion is short-lived, however, and the number returns to 23,000 by January 2021 and then drops away, falling below 10,000 per day by early June 2021. Society and the economy are now fully derestricted, and the rate of new cases drops away to below 100 per day by the beginning of September 2021.

Assuming an initial infection fatality ratio of 500 in 100,000, there will be a further 34,800 deaths between 2 December 2020 and the end of the epidemic. Taking the average life expectancy of Covid-19 victims to be 7.5 years,[58] this implies the loss of 261,000 years of life. This equates to 6,200 average lives lost, based on the UK's population-average life expectancy of about 42 years.

[58] Thomas, P., Is the cost of another lockdown too high? *The Spectator* (8 November 2020) https://www.spectator.co.uk/article/is-the-cost-of-another-lockdown-too-high-

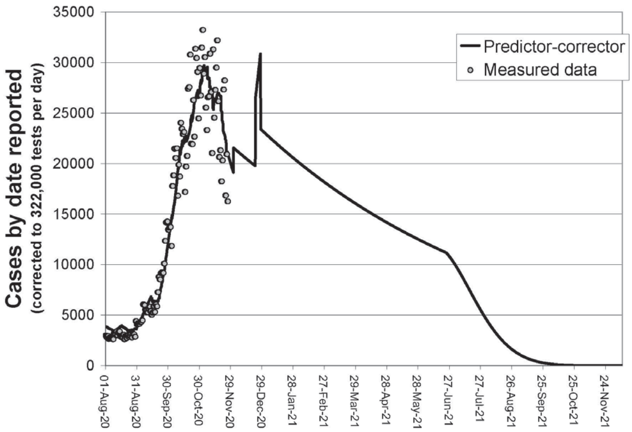

Figure 25. Cases by Date Reported for Scenario 1, §11.1.

12.2 Scenario 2. The SDI is increased to 1.5 on 2 December 2020 maintained at that value, except for the 5 days at Christmas, until 31 March 2021, raised to 2.0 on 1 April 2021 and finally raised to 3.0 (all restrictions removed) on 1 June 2021

The idealized Scenario 1 uses the SDI as a dependent variable that is adjusted continuously to ensure that the R-rate stays just below 1.0 at all times. In Scenario 2, by contrast, the SDI takes its natural role as an independent variable, which the government can, with the consent of the people, control in order to limit the epidemic.

In a procedure very similar to that adopted for Scenario 1, the SDI is assumed to take the value, 1.18, recorded for 26 November 2020 for the days between that date until the last day of lockdown on 1 December 2020.

In practice it would be impossible to vary the SDI continuously because that would require precise changes to restrictions to be made each day as well as their immediate adoption by the public. In reality a small number of clearly understood steps in the level of restrictions is required. Accordingly, in this more realistic second scenario, the SDI is set to the following discrete values:

- 1.5 from 2 December to 22 December 2020
- 1.86 from 23 December to 27 December 2020 (the 5 Christmas days)
- 1.5 from 28 December 2020 to 31 March 2021
- 2.0 from 1 April to 31 May 2021
- 3.0 from 1 June 2021 onwards (complete derestriction).

Excepting the short relaxation over the Christmas period, this corresponds to 3 discrete régimes of restrictions over 7 months (see Figure 26). Comparing Figures 23 and 26, the latter may be interpreted as a discrete approximation to the former. An SDI of 1.5 corresponds roughly to the level seen when the country was just entering the first set of Tier restrictions on 17 October 2020.

The predicted active infections are graphed against time in Figure 27, which shows a gradual rise in cases over the first three weeks after lockdown, up to 630,000, roughly the level of infections in the final week of November 2020. The additional transmission over the Christmas week causes the number of active cases to rise to a peak of 740,000. Cases stay at similarly elevated levels throughout

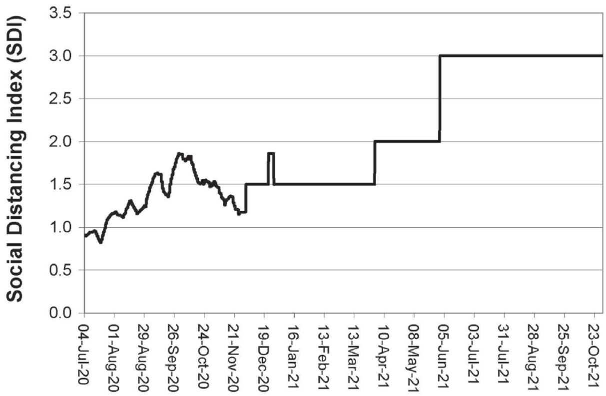

Figure 26. Social distancing index over time, Scenario 2, §11.2.

January 2021 and then begin their long and uninterrupted fall to 160,000 by the end of March 2021. The fall in active cases continues in April and May 2021 despite the further relaxation of the SDI to a value of 2.0 on 1 April 2021. There is a very small uptick when all restrictions are abandoned at the beginning of June 2021 and the SDI rises to 3.0, but active infections continue to decline and drop to below 100 across the country by the middle of October 2021.

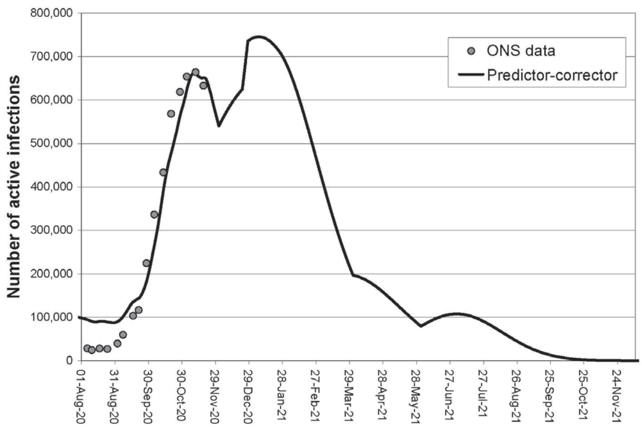

Figure 27. Number of active infections, Scenario 2, §11.2.

Figure 28 shows the Cases by Date Reported against time. The increase in the SDI causes a rise in daily cases during the first three weeks of December 2020, and the Christmas relaxation leads to a spike that reaches 38,000 new cases per day at the end of December 2020, about 5,000 more than the peaks at the end of November 2020. But then the new cases return to 31,000 and, from mid-January 2021 onwards, they begin a continuing fall, dropping to 5,000 a day by the end of March 2021. The two relaxations of restrictions cause minor upticks, but new cases per day remain below 5,000 a day from mid-April 2021 onwards. There are fewer than 100 new cases per day by late August 2021.

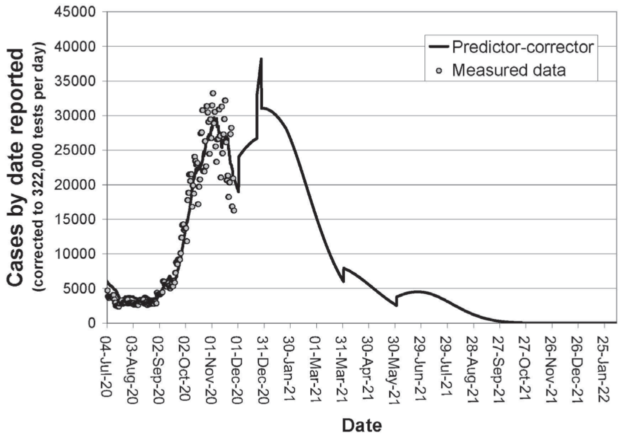

Figure 28. Cases by Date Reported, Scenario 2, §11.2.

The effect on R-rate is shown in Figure 29. It rises a small way above unity as the country comes out of lockdown, but immediately starts to fall as the number of people who have recovered increases and the number of susceptible individuals falls. The Christmas week produces a short-lived pulse, but the R-rate returns quickly to a value marginally above 1.0 and then resumes its steady fall. The effect of the vaccination programme is detectable by the end of January 2021, when the rate of drop-off increases.

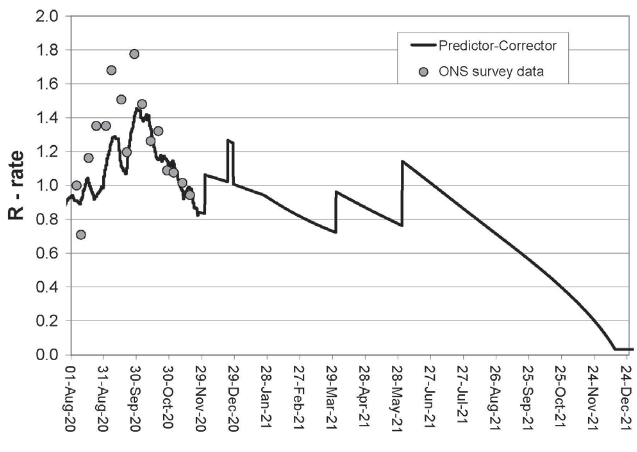

Figure 29. R-rate versus time, Scenario 2, §11.2.

The R-rate falls to 0.7 by the beginning of April 2021. The relaxation of restrictions at the start of April causes it to increase temporarily, but it stays below 1.0 and begins to fall immediately afterwards once again. The final opening up at the beginning of June 2021 causes an increase in the R-rate, but it stays below 1.0. It then starts to fall away steadily and has effectively reached zero by the end of the year.

An initial infection fatality ratio of 500 in 100,000 suggests that there will be a further 35,500 deaths from 2 December 2020 to the end of the epidemic. With an average life expectancy for Covid-19 taken as 7.5 years, this implies the loss of 266,000 years of life. This equates to 6,300 average lives lost, which is roughly the same as under Scenario 1.

Figure 30 compares the death rates for the two scenarios as time proceeds. The recorded figures for daily deaths are also shown. Scenario 2 has rather more deaths between the end of December 2020 and the end of February 2021, but then balances this with fewer later, with the Scenario 2 death rate falling to less than half that of Scenario 1 by the end of March 2021.

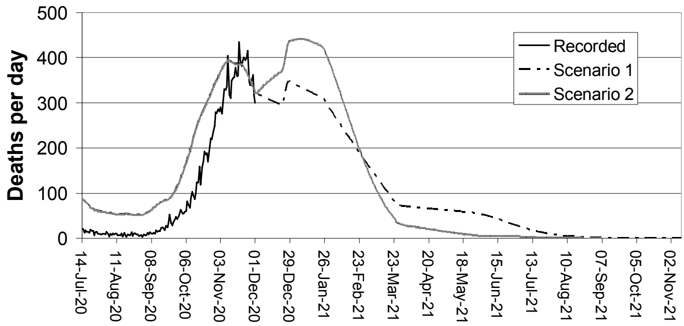

Figure 30. Deaths per day: Scenarios 1 and 2 compared against each other and with recorded daily deaths

13. Discussion

Covid-19 is an unpleasant disease with an infection fatality rate for the general population of roughly 500 in 100,000. This is about 20 times higher than the infection fatality rate estimated for English sufferers from symptomatic influenza of the A/H1N1pdm09 type, or "swine flu", during the pandemic of 2009.[59,60] The chance of dying after being infected with SARS-CoV-2 varies by three orders of magnitude with age in adults and is mainly concentrated in those who are over 65, especially those already afflicted with serious chronic health conditions. The risk of death from Covid-19 for those under the age of 44 is very low. A way for individuals of all ages to gain an intuitive understanding of their personal level of risk should they contract the virus is provided by the temporary age burden (TAB) method. The threat to the person contracting Covid-19 is the same as the extra hazard that he or she would have to face from living normally and infection-free for the next year, but temporarily six to nine years older.

If the problem were simply to reduce the spread of Covid-19 and minimize the associated deaths, the solution would be simple in principle. A strategy of suppressing the virus would be adopted, so that the nation would be kept in permanent lockdown until everyone had been inoculated following the development and roll-out of safe and effective vaccines. This is

[59] Donaldson, L.J. et al., Mortality from pandemic A/H1N1 2009 influenza in England: public health surveillance study. *BMJ* **339** (2009) b5213.

[60] The outbreak was part of a pandemic, but the consequences were much less serious than originally predicted because many older people were already immune to it. See https://www.nhs.uk/conditions/swine-flu/

basically the course of action urged on the government by its medical advisers since the start of the crisis, summarized in the injunction to keep the R-rate below 1.0. But such a policy ignores the very heavy costs that must accompany such a policy's implementation. It has been shown previously that pursuing a lockdown for months risks sacrificing more life immediately and in the following few years than could possibly be lost to the coronavirus.[26,61]

The danger from the outset was that the restrictions imposed by the government would inflict severe damage on the nation's economy and its society. The harmful effects have become now painfully apparent, with the economy experiencing its deepest recession for hundreds of years, businesses having been bankrupted, hundreds of thousands of people losing their jobs, the national debt climbing to unprecedented levels and children losing months of schooling.

In reality the country is faced with a difficult systems problem where trade-offs need to be made between benefits of different kinds. Some have suggested that the government faces a simple binary choice: either lockdown and suppress or "let the virus rip", because any spread will be exponential and an unchecked exponential will lead to unlimited runaway. In reality the growth will at no stage be truly exponential and the growth is increasingly resisted to the point where it comes up against a hard limit when population ("herd") immunity is reached, at which point the epidemic will recede.

A suspicion of herd immunity seems to have developed in the media, possibly because it is regarded as synonymous with two thirds of the population being infected, a large number. But the all-important SDI determines the fraction of the population that needs to be immune before the epidemic goes into decline. While two thirds would be needed if the SDI stayed at its original value of 3.0, when the SDI is lowered to 1.5, its value just before England's first Tier restrictions were introduced in October 2020, then only 33% of the population need be immune for the epidemic to be forced into retreat. This is not far above the 28% of the population of England that is calculated to have become immune to the coronavirus by the end of November 2020.

The SDI is what the government, acting with the consent of the people, has the power to control. But before something can be effectively controlled, it needs to be measured. As discussed in §9, the government seems to have had no up-to-date measurement of either the SDI or the R-rate to inform its decision to impose a second lockdown on England on 5 November 2020. Since England's R-rate was either approaching or below 1.0 at the time, it is possible that the Prime Minister would have acted differently if he had been aware of the fact.

It is clear that a robust and rapid measurement system is needed by the country's decision-makers as a matter of urgency if an optimal course is to be steered over the coming months before the nation is free of the coronavirus. The measurement system described in detail in this paper, and in summary previously,[62] would allow timely measurements of both the social distancing index and the R-rate.

The input to the predictor–corrector is the daily figure for Cases by Date Reported, which the Government puts out as the headline figure by each day. It is important that this be corrected to standard figures for the numbers of tests carried out in Pillars 1 and 2. The reference figures

[61] Thomas, P., Is the cost of another lockdown too high? *The Spectator* (8 November 2020) https://www.spectator.co.uk/article/is-the-cost-of-another-lockdown-too-high-

[62] Thomas, P., 2020, The path between herd immunity and lockdown, *The Spectator*, 24 October, https://www.spectator.co.uk/article/the-path-between-herd-immunity-and-lockdown#

used in this paper are the 71,749 tests under Pillar 1 and 250,067 tests under Pillar 2 that were reported on 13 November 2020 for England. Expected values for the number of Cases by Date Reported for other test numbers may be derived using the method set out in detail in Appendix D. The correction exercise uses data from the NHS England COVID-19 Daily Situation Report and from NHS Test and Trace Data Tables published weekly.

Given that health is a devolved area of policy, access to different databases for the four countries would be needed to find all the information required to cover each of the four home countries. Each would require a slightly different approach. Given the urgency to produce useful results, it was decided to concentrate the analysis on England and to compose scenarios for that country. England makes up 84% of the population of the United Kingdom and obviously has high policy-making relevance. Nevertheless, the approach could in principle be extended to each of the countries making up the UK and to regions within those countries, subject to the data being available.

The corrected Cases by Date Reported provide perspective on the severity of the second wave. The peak of infections experienced in the second wave is shown to be only half that seen in the first wave once the figures have been corrected. The correction is also able to explain how some daily figures for new cases, which might have been classified as outliers previously, actually lie well within levels of expected variation.

The overall measurement system has been shown to provide an accurate measurement each day of:

- the number of active infections,
- the R-rate and
- the SDI

valid for 4 days before. This is, on average, ten days faster than equivalent measurements based on ONS survey data can be generated and the accuracy has been shown to be comparable.

The predictor–corrector measurement method operates in close to real time and thus allows policymakers a new freedom. They will be able, for the first time, to choose a setpoint for the SDI and then judge within 4 days how well people are conforming to the new requirement. They will also be able to see the number of people infected and the R-rate at the same time.

The measurement system makes proper allowance for the T-cell immunity level that has recently been quantified. Thanks to work by Public Heath England and its university collaborators, it is now known that 12.9% of the population has prior immunity to the coronavirus. This is a key finding, vital to our understanding of the epidemic and prediction of its course.

Using this figure for T-cell immunity within the corrector model within the measurement system allows the number of people with an active infection to be measured accurately. The model incorporating this level of T-cell immunity receives substantial validation from its agreement with ONS survey estimates on how active infections have evolved over time, as shown in Figure 17.

The Pfizer-BioNTec vaccine was approved by the Medicines and Healthcare products Regulatory Agency (MHRA) for public use in the UK on 2 December 2020.[63] The Moderna and AstraZeneca/Oxford vaccines are expected to be approved in the near future. A model for mass

---

[63] https://www.nhs.uk/conditions/coronavirus-covid-19/coronavirus-vaccination/coronavirus-vaccine/?gclid=EAIaIQobChMI44Gfh_Cx7QIV1u7tCh0oSgFZEAAYASAAEgIzlvD_BwE (note added in proof).

vaccination has been developed in Appendices F and G and added to the corrector system. This allows assessments to be made of different time-profiles for SDI setpoints over the coming year.

Scenario 1 assumes that the setpoint for the SDI is varied each day so as to maintain the R-rate at 0.95, except at Christmas, when the SDI will experience a temporary pulse to take it to 1.86 for 5 days, this being the maximum level observed in England since the first lockdown. The scenario conforms to current government policy, which requires the R-rate to be below 1.0 except for the 5 days of Christmas.

Government can stay within its rules by raising the setpoint for the SDI continuously through relaxing restrictions each day to the beginning of June 2021, at which point the economy and society will be free of all restrictions. Apart from temporary rises caused by the Christmas relaxation of rules, both active infections and Cases by Date Reported will fall continuously and the epidemic will be essentially over by the autumn of 2021. No Covid-19 deaths are likely after the middle of September 2021.

The daily variation in the SDI envisaged in Scenario 1 would be a practical impossibility, of course. Therefore Scenario 2 was designed to have three discrete levels of the SDI: 1.5 on 2 December 2020, 2.0 on 1 April 2021 and 3.0 on 1 June 2021, when all restrictions are lifted. This trajectory for the SDI may be regarded as a discrete approximation to the continuous trace seen under Scenario 1.

The intial value of the SDI, 1.5, under this scenario, is the value it had just before the first set of Tier restrictions were imposed on 17 October 2020. While by no means implying freedom from restrictions, this level of SDI would give the economy and society a better chance of recovery than its likely value under the stricter new Tier restrictions.

The daily deaths under Scenario 2 start off slightly higher than under Scenario 1, but they fall quicker after the end of March 2021 so that they reach an approximate balance by the end of the epidemic in the autumn of 2021.

14. Conclusions

A new predictor–corrector measurement tool has been described that is able to provide a good estimate of the social distancing index, also known as the basic reproduction number, across England on a daily basis. The system also measures the number of active infections in the country and the R-rate.

The measurement is based on the number of Cases by Date Reported supplied each day on the UK government's Coronavirus Dashboard. While the Cases by Date Reported are found theoretically to lag on the Cases by Specimen Date by 2 days, the 5-day delay in assembling the data for the latter means that Cases by Date Reported are the most timely indicator of the number of new cases daily.

The measurements of the social distancing index, the R-rate and the number of current infections are continuous and valid for a date 4 days earlier. These measurements are available 10 days, on average, before the equivalent measurements based on data from the Office of National Statistics Coronavirus Infection Survey Pilot. The predictor–corrector system has been shown to give comparable accuracy to spot figures calculated from ONS survey data.

The measurement system has shown that England's R-rate was either approaching or below 1.0 at the time the country entered its second lockdown on 5 November 2020. The action of the government in imposing the second English lockdown suggests it might not have been aware of this.

The capability of the new measurement system to provide timely feedback to decision-makers allows them, for the first time, freedom to control the SDI to a setpoint of their choosing. Policymakers need no longer see their choices of Covid-19 strategies as binary: either allow the virus to spread, aiming to achieve herd immunity as quickly as possible or else try to hold it off until the country has been vaccinated. The predictor–corrector measurement system allows a third way, whereby the SDI is controlled to a series of setpoints selected to minimize the overall harm to the country's health and economy.

The proper functioning of the measurement has required the development of a compensating factor to be applied to the Cases by Date Reported to correct for the large disparities in the numbers of tests carried out each day. This allows comparability for the first time between daily numbers of cases on different dates. *Inter alia*, it shows that despite being apparently 8 times bigger, peak cases in the second wave in October–November 2020 were only half the height seen in the first wave, March–April 2020, on a like-for-like basis.

Essential to the accuracy of the predictor–corrector measurement system is the need to account for the level of prior immunity in the population due to T-cells. Recent research that has shown roughly 1 in 8 people to carry such immunity has been confirmed by the validation of the predictor–corrector model against ONS survey data on the number of active infections over time in England.

Scenarios for mass vaccination of the English population from December 2020 to the end of the epidemic in England have been examined. An idealization of the government's declared policy of keeping the R-rate below 1.0 was considered in the first scenario. This could be achieved while attempting to minimize harm to the economy by loosening restrictions continuously from December 2020 to the beginning of June 2021. The second scenario is effectively an attempt at practical implementation of the first, effected in three steps. The extra immunity that the population possesses allows an immediate return to the level of restrictions seen just before the first Tiers was imposed. A further relaxation occurs at the start of April and all restrictions are removed at the beginning of June 2021. In both scenarios about 35,000 lives are estimated to be lost between 2 December 2020 and the end of the epidemic in England, equivalent to about 6,200 average UK lives. However the second scenario would reduce the early harm to society and the economy.

Appendix A. Development of the temporary age burden (TAB) method

People's chances of dying from other causes are found to rise approximately exponentially with age, as observed by Spiegelhalter[9] and documented in life tables. The estimated probability $m_X(x)$ of dying at age x, given that the person has reached exact age x, may thus be written:

$$m_X(x) = be^{ax} \qquad \text{(A.1)}$$

with a and b constants. Hence, after taking logarithms,

$$\ln\left(m_X(x)\right) = \ln ax + b. \qquad \text{(A.2)}$$

From equation (A.1), the ratio r of hazard rates, $r = m_X(x_2)/m_X(x_1)$ at age x_1 and a later age x_2 will be

$$r = \frac{m_X(x_2)}{m_X(x_1)} = \mathrm{e}^{a(x_2 - x_1)}. \qquad \text{(A.3)}$$

Thus, after again taking logarithms,

$$x_2 - x_1 = \Delta x = \frac{\ln r}{a}. \qquad \text{(A.4)}$$

For a doubling of the hazard rate, $r = 2$. Thus the additional age needed before the hazard rate is doubled is

$$\Delta x = \frac{0.6931}{a}. \qquad \text{(A.5)}$$

A very good match to the latest UK life tables[11] may be found by taking the ranges: 20–59, 60–70 and 71–100 and fitting a linear curve to the resultant log–linear graphs (Figures A.1, A.2 and A.3). As seen by eye and measured by the square of the correlation coefficient, the matches are very good.

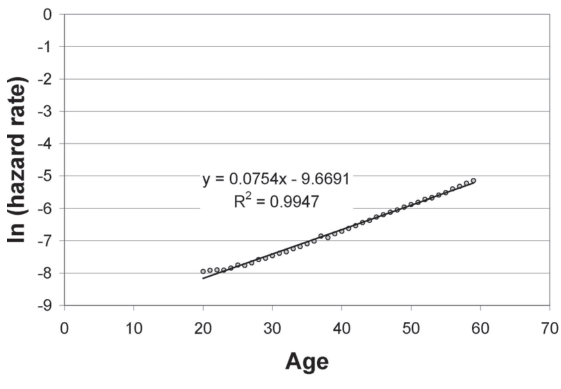

Figure A.1. Logarithm of hazard rate for ages 20–59 (here R^2 is the correlation coefficient squared).

The values of Δx may be found by substituting the value of a found in each of the three cases into equation (A.5) and the results are shown in Table A.1. So, for someone aged between 20 and 59, the risk of dying from Covid-19 having just caught it equals the extra risk of living normally and Covid-free for the next year but being 9 years older throughout the period. The extra age is then removed entirely after the year is up.

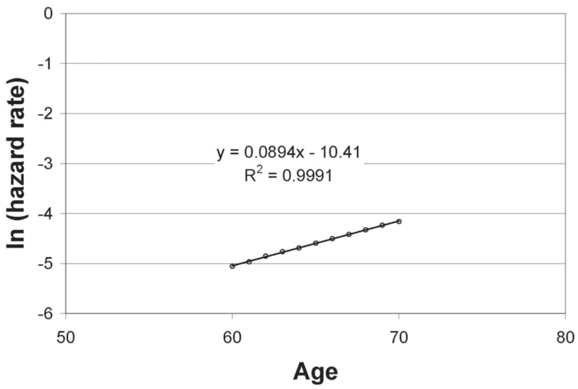

Figure A.2. Logarithm of hazard rate for ages 60–70 (here R^2 is the correlation coefficient squared).

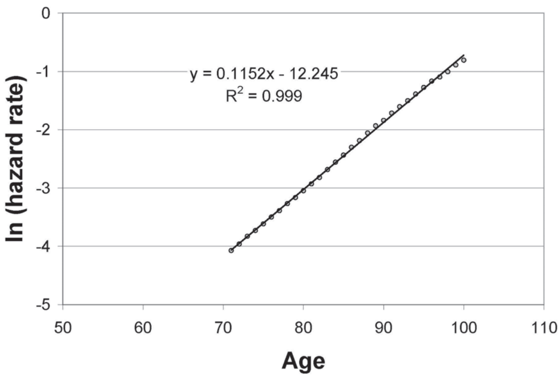

Figure A.3. Logarithm of hazard rate for ages 71–100 (here R^2 is the correlation coefficient squared).

Table A.1. Additional age burden, Δx, needed to double the hazard rate for those aged 20 years and above.

Age range/years	Additional age, Δx / years, needed to double the hazard rate
20–59	9.2
60–70	7.8
71+	6.0

For someone between 60 and 70, the extra burden is 8 years. Those catching Covid over the age of 70 will need to add on just 6 years to their age for a year, reflecting the fact that the increase in hazard rate with age has speeded up.

There are two advantages to this extension of Spiegelhalter's observation as far as communication with the public is concerned. First, most people have little intuitive understanding of probability and risk. This is not surprising, as the subject is inherently difficult to comprehend. Hence many may well misunderstand the (entirely fair) statement that the chance of dying from Covid-19 after having caught it is roughly the same as faced by just living through the next year and the normal hazards it brings. Indeed, Spiegelhalter notes that such misunderstandings have happened (see 4th last paragraph of the section, *Infection fatality risks in terms of normal risks*[9]). In contrast, people understand very well that older people have a higher chance of dying from natural causes. This comes directly from their own observation of the world around them. Presenting the extra risk in terms of being older for a year stands a better chance of getting the message through.

Secondly, the TAB presentational method gets over, at least to some extent, the reservation expressed by Spiegelhalter in the last paragraph of the section, *Infection fatality risks in terms of normal risks*:[9]

> Importantly, all the risks quoted are the average (mean) risks for people of the relevant age but are not the risks of the average person. This is because, both for covid-19 and in normal circumstances, much of the risk is held by people who are already chronically ill. So for the large majority of healthy people, their risks of either dying from covid-19, or dying of something else, are much lower than those quoted here.

But, when given the guidance in terms of the burden of living at a greater age for a year, everyone will automatically and quite naturally consider how he or she will be faring in a few years' time. It would be reasonable for healthy people to see themselves as part of a cohort likely to continue in good health for some time to come and weigh up the effect of the extra age optimistically. Meanwhile, those in poorer health might see themselves as part of a struggling cohort and come to a more pessimistic conclusion. By considering his or her personal health situation, each person will come up with an interpretation of the information received that is likely to be broadly correct.

Appendix B. Correcting the data for the computer error for England, 25 September 2020 to 4 October 2020

A file overload occurred in transferring data from testing centres to the UK government's coronavirus dashboard,[64] so that 15,841 cases of Covid-19 in England between 25 September 2020 and 2 October 2020 were left out of England's (and the UK's) daily case figures. The extra cases were then added to the figures for Cases by Date Reported for 3 October 2020 and 4 October 2020. These corrected the cumulative total, but gave an unrealistic picture of the development of the second wave.

The figure for new cases is known for 24 September 2020: $c_0 = 6634$, where c_k stands for new cases on date, k. The corresponding figure is also known 11 days later on 5 October 2020: $c_{11} = 12594$. The requirement is to assign figures to c_k, $k = 2, 3, \ldots, 10$.

Let the daily figures currently published be $c'_k, k = 2, 3, \ldots, 10$. While each is individually unrealistic, it is known from a comment on the website[64] that

$$\sum_{k=1}^{8} c_k = \sum_{k=1}^{8} c'_k + 15841 = 66631 \tag{B.1}$$

[64] Coronavirus (COVID-19) in the UK, Cases by date reported (locate headings: Daily, Cumulative, Data, About; click on "About") https://coronavirus.data.gov.uk/cases

since adding the website figures gives $\sum_{k=1}^{8} c'_k = 50790$. This implies that the average number of daily cases between 25 September and 2 October 2020 was 8329. Moreover, since the additional 15,841 cases were added in to the figures for October 3 and October 4 2020, c_9 and c_{10}, we have:

$$\sum_{k=1}^{10} c_k = \sum_{k=1}^{10} c'_k = 86623. \tag{B.2}$$

But

$$\sum_{k=1}^{10} c_k = \sum_{k=1}^{8} c_k + \sum_{k=9}^{10} c_k = 66631 + \sum_{k=9}^{10} c_k. \tag{B.3}$$

Combining equations (B.2) and (B.3) gives

$$\sum_{k=9}^{10} c_k = 86623 - 66631 = 19992. \tag{B.4}$$

This implies that the average figure for daily cases was 9996 between 3 and 4 October 2020.

Equations (B.1) and (B.4) act as constraints on the allocation of cases from 25 September to 2 October 2020 and from 3 October to 4 October 2020 respectively.

Two simple types models may be used for the allocation of the cases: linear and exponential.

B.1 Linear models

Interval 25 September 2020 to 2 October 2020

A linear model implies that the cases will rise by the same amount, a, each day, so that

$$c_k = c_0 + ka \quad \text{for } k = 1, 2, ..., 8 \tag{B.5}$$

subject to the constraint of equation (B.1). But

$$\sum_{k=1}^{8} c_k = \sum_{k=1}^{8} \left(c_0 + ka\right) = 8c_0 + a\sum_{k=1}^{8} k = 8c_0 + a\frac{8\left(8+1\right)}{2} = 8c_0 + 36a. \tag{B.6}$$

Hence

$$a = \frac{66631 - 8 \times 6634}{36} = 376.6. \tag{B.7}$$

Interval 3 October 2020 to 4 October 2020

Now the number of cases reported for 5 October 2020 may be used, so that

$$c_k = c_{11} - \left(11 - k\right)b \quad \text{for } k = 9, 10 \tag{B.8}$$

so that

$$c_9 + c_{10} = 2c_{11} - 3b. \tag{B.9}$$

Combining equations (B.4) and (B.9) gives

$$b = \frac{2 \times 12594 - 19992}{3} = 1732. \tag{B.10}$$

B.2 Exponential models

Interval 25 September 2020 to 2 October 2020

By the exponential model:

$$c_k = c_0 e^{a_1 k} \quad 1 \le k \le 8 \tag{B.11}$$

where a_1 is a constant. Applying the constraint of equation (B.1)

$$\sum_{k=1}^{8} c_k = c_0 \sum_{k=1}^{8} e^{a_1 k} = 66631 \tag{B.12}$$

so that

$$\sum_{k=1}^{8} e^{a_1 k} = \frac{66,631}{6,634} = 10.044. \tag{B.13}$$

This may be solved numerically to give $a_1 = 0.049153$.

Interval 3 October 2020 to 4 October 2020

Using the number of cases reported for 5 October 2020 as a basis

$$c_k = c_{11} e^{-b_1 k}, \quad k = 9,10. \tag{B.14}$$

Thus

$$c_9 + c_{10} = c_{11}\left(e^{-b_1} + e^{-2b_1}\right). \tag{B.15}$$

Combining this with the constraint of equation (B.4) gives

$$e^{-b_1} + e^{-2b_1} = \frac{19992}{12594} = 1.5874. \tag{B.16}$$

A numerical solution gives $b_1 = 0.156051$.

B.3 Mixed linear and exponential models

It is possible to combine the two models according to

$$c_k = \begin{cases} f\left(c_0 + ka\right) + \left(1-f\right)c_0 e^{a_1 k} & 1 \le k \le 8 \\ f\left(c_{11} + \left(11-k\right)b\right) + \left(1-f\right)c_{11} e^{-b_1 k} & k = 9,10 \end{cases} \quad 0 \le f \le 1 \tag{B.17}$$

while still conforming to the constraints of equations (B.1) and (B.4), as can be seen by carrying out the appropriate summations.

All the allocation methods, linear, exponential and mixed, provide an averaging and smoothing effect on the data. The values of c_k produced by all the methods lie very close to each other. The linear formulation produces a slightly smaller dip between c_8 and c_9 and is slightly simpler. It is therefore marginally to be preferred will therefore be adopted.

A comparison of the daily cases presented on the coronavirus dashboard and the linear allocation method is shown in Figure B.1. The smoothed version is likely to be a lot more realistic.

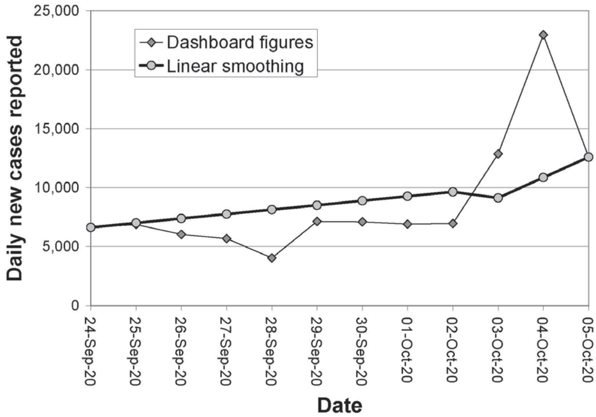

Figure B.1. Comparing the data presented on the Coronavirus Dashboard with smoothed figures that preserve the total cases reported over the intervals 25 September to 2 October and 3 October to 4 October.

Appendix C. Expected value and variance of the estimate of *R* found from ONS surveys

This estimate of R relies on the numbers of active infections in the country at two different dates t_b and t being found by sampling. This introduces an element of chance—a different sample would probably have given different values—thereby rendering both random variables. The calculated value of R will therefore also be a random variable. The three random variables will be denoted: $\widehat{n}(t_b), \widehat{n}(t)$ and $\widehat{R}$, and each will have a mean value and a variance. ONS provides the upper and lower figures for a 95% confidence interval, allowing an approximate calculation of the standard deviation as the upper minus the lower divided by 3.92, corresponding to an assumed normal distribution.

C.1 Expected value

Applying equation (14), the basic reproduction number, now regarded as a random variable, may be written:

$$\widehat{R} \approx 1 + \frac{\tau_{\text{inf}}}{t - t_0} \ln \frac{\widehat{n}(t)}{\widehat{n}(t_b)}. \tag{C.1}$$

Applying the expectation operator $E(.)$ to equation (C.1) reveals that the mean or expected value of $\widehat{R}$ will be:

$$E(\widehat{R}) \approx E(1) + E\left(\frac{\tau_{\text{inf}}}{t - t_b} \ln \frac{\widehat{n}(t)}{\widehat{n}(t_b)}\right) = 1 + \frac{\tau_{\text{inf}}}{t - t_b} E\left(\ln \frac{\widehat{n}(t)}{\widehat{n}(t_b)}\right). \tag{C.2}$$

In general, if $Y = h(X_1, X_2)$ is a nonlinear function of two random variables X_1 and X_2 with means μ_{X1} and μ_{X2} and variances σ^2_{X1} and σ^2_{X2}, then it can be shown that[65]

[65] E.g., §4.6 Approximate methods. In: Rice, J. A., *Mathematical Statistics and Data Analysis*. Andover: Cengage Learning (2007).

$$E(Y) \approx h(\mu_{X1}, \mu_{X2}) + \frac{1}{2}\left(\frac{\partial^2 h(\mu_{X1}, \mu_{X2})}{\partial X_1^2} \sigma_{X1}^2 + 2\frac{\partial^2 h(\mu_{X1}, \mu_{X2})}{\partial X_1 \partial X_2} \sigma_{X1,X2} + \frac{\partial^2 h(\mu_{X1}, \mu_{X2})}{\partial X_2^2} \sigma_{X2}^2 \right) \quad \text{(C.3)}$$

where $\sigma_{X1,X2} = E((X_1 - \mu_{X1})(X_2 - \mu_{X2}))$ is the covariance between X_1 and X_2. In the case under consideration, the function, $h(X_1, X_2)$, is given by

$$h(X_1, X_2) = \ln U \quad \text{(C.4)}$$

where

$$U = \frac{X_2}{X_1}. \quad \text{(C.5)}$$

Differentiation yields:

$$\begin{gathered} \frac{\mathrm{d}h}{\mathrm{d}U} = \frac{1}{U} = \frac{X_1}{X_2}; \quad \frac{\partial U}{\partial X_1} = \frac{-X_2}{X_1^2}; \quad \frac{\partial U}{\partial X_2} = \frac{1}{X_1} \\ \frac{\partial h}{\partial X_1} = \frac{\mathrm{d}h}{\mathrm{d}U}\frac{\partial U}{\partial X_1} = \frac{-1}{X_1}; \quad \frac{\partial h}{\partial X_2} = \frac{\mathrm{d}h}{\mathrm{d}U}\frac{\partial U}{\partial X_2} = \frac{1}{X_2} \\ \frac{\partial^2 h}{\partial X_1^2} = \frac{1}{X_1^2}; \quad \frac{\partial^2 h}{\partial X_2^2} = \frac{-1}{X_2^2}; \quad \frac{\partial^2 h}{\partial X_1 \partial X_2} = 0. \end{gathered} \quad \text{(C.6)}$$

The fact that $\frac{\partial^2 h}{\partial X_1 \partial X_2} = 0$ means that the covariance term in equation (C.3) will disappear, so that

$$E\left(\ln \frac{X_2}{X_1}\right) \approx \ln \frac{\mu_{X2}}{\mu_{X1}} + \frac{1}{2}\left(\frac{1}{\mu_{X1}^2}\sigma_{X1}^2 - \frac{1}{\mu_{X2}^2}\sigma_{X2}^2 \right). \quad \text{(C.7)}$$

In the case at hand,

$$\begin{aligned} \mu_{X1} &= E(\hat{n}(t_b)) \\ \mu_{X2} &= E(\hat{n}(t)) \end{aligned} \quad \text{(C.8)}$$

while

$$\begin{aligned} \sigma_{X1}^2 &= \mathrm{var}(\hat{n}(t_b)) \\ \sigma_{X1}^2 &= \mathrm{var}(\hat{n}(t)). \end{aligned} \quad \text{(C.9)}$$

C.2 Variance

Applying the variance operator, var(.), to equation (C.1) gives the variance of $\hat{R}$ as:

$$\mathrm{var}(\hat{R}) \approx \left(\frac{\tau_{\mathrm{inf}}}{t - t_b} \right)^2 \mathrm{var}\left(\ln \frac{\hat{n}(t)}{\hat{n}(t_b)} \right). \quad \text{(C.10)}$$

For the general nonlinear function $Y = h(X_1,X_2)$ the variance will be approximated[66] by:

$$\mathrm{var}(Y) \approx \left(\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_1}\right)^2 \sigma_{X1}^2 + 2\rho_{X1,X2}\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_1}\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_2}\sigma_{X1}\sigma_{X2} + \left(\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_2}\right)^2 \sigma_{X2}^2 \tag{C.11}$$

where $\rho_{X1,X2}$ is the correlation coefficient between X_1 and X_2. In the case in hand, the ratio, $\hat{n}(t)/\hat{n}(t_b)$, may be written so as to bring out the increment, $\Delta\hat{n}(t) = \hat{n}(t) - \hat{n}(t_b)$:

$$\frac{\hat{n}(t)}{\hat{n}(t_b)} = 1 + \frac{\Delta\hat{n}(t)}{\hat{n}(t_b)}. \tag{C.12}$$

Now a series expansion of $\ln(1 + x)$ gives:

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \ldots \approx x \qquad \text{for small } x. \tag{C.13}$$

The match between $\ln(1 + x)$ and x is displayed for $0.5 \le x < 1$ in Figure C.1, from which it is clear that the match is very good for $0.8 \le x \le 1.2$. but does not deteriorate dramatically over the full range. Hence the approximation $x \approx \ln(1 + x)$ may be used in the generation of variances, with the understanding that it can be expected to give less accurate answers when either $x <\sim 0.8$ or $x >\sim 1.2$.

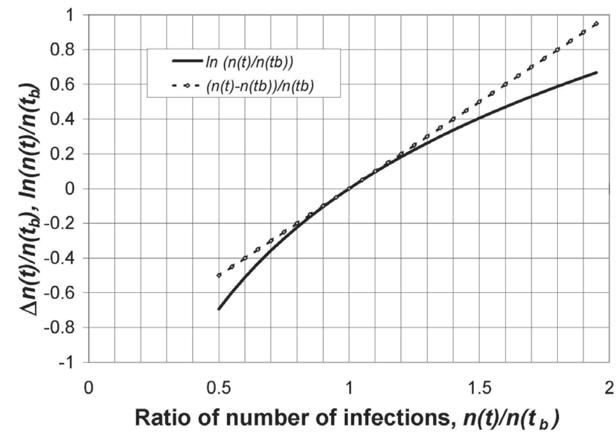

Figure C.1. Comparison of the functions $(n(t) - n(t_b))/n(t_b)$ and $\ln(n(t)/n(t_b))$.

When

$$h(X_1,X_2) = \ln\left(1 + \frac{X_2}{X_1}\right) \approx \frac{X_2}{X_1} \tag{C.14}$$

[66] §4.6 Approximate methods. In: Rice, J. A., *Mathematical Statistics and Data Analysis*. Andover: Cengage Learning (2007).

differentiation yields:

$$\frac{\partial h}{\partial X_1} = \frac{-X_2}{X_1^2}; \qquad \frac{\partial h}{\partial X_2} = \frac{1}{X_1}. \tag{C.15}$$

Applying to equation (C.11) gives:

$$\begin{aligned} \operatorname{var}\left(\frac{X_2}{X_1}\right) &\approx \left(\frac{\mu_{X2}}{\mu_{X1}^2}\right)^2 \sigma_{X1}^2 + \left(\frac{1}{\mu_{X1}}\right)^2 \sigma_{X2}^2 - 2\rho_{X1,X2}\frac{\mu_{X2}}{\mu_{X1}^3}\sigma_{X1}\sigma_{X2} \approx \\ &\left(\frac{\mu_{X2}}{\mu_{X1}}\right)^2 \frac{\sigma_{X1}^2}{\mu_{X1}^2} + \left(\frac{\mu_{X2}}{\mu_{X1}}\right)^2 \frac{\sigma_{X2}^2}{\mu_{X2}^2} - 2\rho_{X1,X2}\frac{\mu_{X2}}{\mu_{X1}}\frac{\sigma_{X1}}{\mu_{X1}}\frac{\sigma_{X2}}{\mu_{X1}}. \end{aligned} \tag{C.16}$$

In the case of equation (C.12), the relevant means and variances, μ_{X1}, μ_{X2}, σ_{X1}^2 and σ_{X2}^2, are given by the following equations:

$$\mu_{X1} = E\left(\hat{n}(t_b)\right) \tag{C.17}$$

$$\mu_{X2} = E\left(\hat{n}(t)\right) - E\left(\hat{n}(t_b)\right) \tag{C.18}$$

$$\sigma_{X1}^2 = \operatorname{var}\left(\hat{n}(t_b)\right) \tag{C.19}$$

$$\begin{aligned} \sigma_{X2}^2 &= \operatorname{var}\left(\hat{n}(t)\right) + \operatorname{var}\left(\hat{n}(t_b)\right) - 2\operatorname{cov}\left(\hat{n}(t), \hat{n}(t_b)\right) = \\ &= \operatorname{var}\left(\hat{n}(t)\right) + \operatorname{var}\left(\hat{n}(t_b)\right) - 2\rho_{\hat{n}(t),\hat{n}(t_b)}\sigma_{\hat{n}(t)}\sigma_{\hat{n}(t_b)}. \end{aligned} \tag{C.20}$$

Plotting $\hat{n}(t)$ against $\hat{n}(t_b)$ when cases are steady or rising shows a strong positive correlation, with $\rho_{\hat{n}(t),\hat{n}(t_b)}$ estimated as 0.99.

The behaviour of the correlation coefficient, $\rho_{X1,X2} = \rho_{\hat{n}(t_b),\Delta\hat{n}(t)}$, as measured by values of $\hat{n}(t)$ and $\hat{n}(t_b)$ is a little more complicated. Figures C.2 and C.3 shows that there is a strong positive correlation when $\Delta\hat{n}(t)$ is increasing, followed by a strong negative correlation when $\Delta\hat{n}(t)$ is decreasing. The associated correlation coefficients are $\rho_{\hat{n}(t_b),\Delta\hat{n}(t)} = 0.87$ and $\rho_{\hat{n}(t_b),\Delta\hat{n}(t)} = -1.0$.

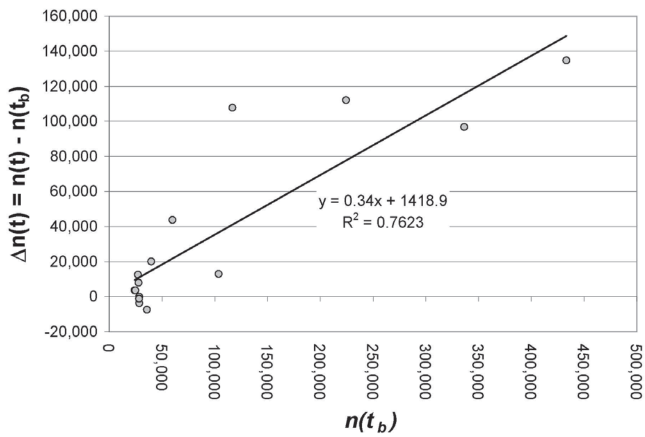

Figure C.2.Correlation between $\Delta n(t)$ and $n(t_b)$ when $\Delta n(t)$is rising (here R^2 is the correlation coefficient squared).

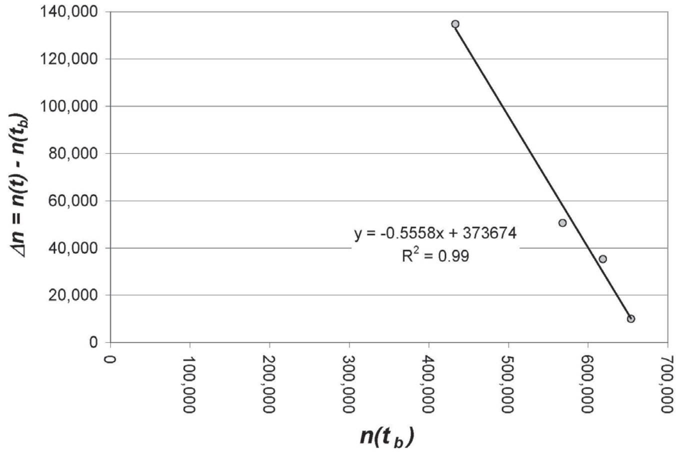

Figure C.3.Correlation between $\Delta n(t)$ and $n(t_b)$ when $\Delta n(t)$ is falling (here R^2 is the correlation coefficient squared).

Appendix D. Modelling the effect of differences in the number of tests on different days

D.1 Introduction

The number of viral tests rose from a few thousand a week at the beginning of the pandemic, to 10,000 a day at the end March 2020, and to over 300,000 per day by the beginning of November 2020. This will clearly affect the numbers of active infections reported.

D.2 Categorization of cases by Pillars

The Cases by Date Reported are categorized using "Pillars". The relevant Pillars are

- Pillar 1, which contains data for NHS swab testing for active viruses in those individuals with a clinical need and, from early April 2020 onwards, health and care workers;
- Pillar 2, which contains data for swab testing for active viruses in the wider population, but only for those who have symptoms that could indicate Covid-19. Pillar 2 information is provided for England from 13 July 2020 onwards.

Pillar 1 began by focusing almost exclusively on NHS patients with a possible medical need,[67] and its extension to include NHS staff started around the beginning of April 2020:[67]

> 1.12 In line with our strategy, the majority of this [Pillar 1] testing has been for those with a medical need—for example, identifying coronavirus sufferers amongst the most unwell people in hospital to allow clinicians to separate them from other vulnerable patients and provide the best possible clinical care; or for investigating specific outbreaks.

[67] Department of Health and Social Care, *COVID-19—Scaling up our testing programmes* (4 April 2020) https://www.gov.uk/government/publications/coronavirus-covid-19-scaling-up-testing-programmes

1.13 Within these numbers, some critical NHS key workers in these hospitals have already been tested for COVID-19, where there has been spare capacity based on the decision of local Trusts.

...

1.15 As of 4th April, nearly 4,500 NHS workers have been tested [out of ~200,000 tests (*ibid.*)] in the pilot stage across five new testing sites.

Let the number of people with a clinical need who selected for testing under Pillar 1 on date t be $w_{11}(t)$ and the number selected from NHS workers under Pillar 1 on the same date be $w_{12}(t)$. Hence the total number of people $w_1(t)$ chosen under Pillar 1 on date t will be the sum:

$$w_1(t) = w_{11}(t) + w_{12}(t). \tag{D.1}$$

Meanwhile the number of tests processed for the general public, classified as Pillar 2, will be denoted $w_2(t)$. These tests under Pillar 2 are generally available only for people experiencing symptoms.

Let the number of positive tests in a day be $M(t)$, which may be regarded as a binomial random number that is the sum of Bernoulli random variables, the value of which will be 1 if the test is positive and 0 otherwise. The number $M(t)$ will be made up of three streams, so that

$$M(t) = M_{11}(t) + M_{12}(t) + M_2(t) \tag{D.2}$$

where $M_{11}(t)$ is the number of patients testing positive on day t, $M_{12}(t)$ is the number of NHS workers testing positive on day t, and $M_2(t)$ is the number of symptomatic members of the general public testing positive on day t. Let the associated probabilities of a positive test be $p_{11}(t)$, $p_{12}(t)$ and $p_2(t)$. Then the expected value of the number of patients testing positive on day t will be:

$$E(M_{11}(t)) = p_{11}(t)w_{11}(t) \tag{D.3}$$

and so on. Moreover,

$$E(M(t)) = E(M_{11}(t)) + E(M_{12}(t)) + E(M_2(t)) \tag{D.4}$$

and hence

$$E(M(t)) = p_{11}(t)w_{11}(t) + p_{12}(t)w_{12}(t) + p_2(t)w_2(t). \tag{D.5}$$

Let $E(M(t))$, $E(M_{11}(t))$,... be estimators for $m, m_{11}, \ldots$, so that:

$$\begin{aligned} m(t) &\approx E(M(t)) \\ m_{11}(t) &\approx E(M_{11}(t)) \\ m_{12}(t) &\approx E(M_{12}(t)) \\ m_2(t) &\approx E(M_2(t)). \end{aligned} \tag{D.6}$$

Combining equations (D.5) and (D.6) gives

$$m(t) \approx m_{11}(t) + m_{12}(t) + m_2(t) \tag{D.7}$$

where

$$\begin{aligned} m_{11}(t) &\approx p_{11}(t)\,w_{11}(t) \\ m_{12}(t) &\approx p_{12}(t)\,w_{12}(t) \\ m_{2}(t) &\approx p_{2}(t)\,w_{2}(t). \end{aligned} \tag{D.8}$$

The total number of positive tests per day $m(t)$ may be approximated by the "Cases by Date Reported" supplied daily on the UK Government's Coronavirus Dashboard. The same dashboard provides the daily number of tests in Pillar 2, $w_2(t)$, and also the daily number of tests in Pillar 1, $w_1(t)$. However, the latter figure is not broken down into $w_{11}(t)$ and $w_{12}(t)$, which therefore need to be estimated.

The number of positive tests for hospital patients in NHS England is given monthly in arrears by the COVID-19 Daily Situation Report spreadsheet.[68] This provides $m_{11}(t)$. The same situation report provides the number of people in NHS England taking sickness absence on any given day due to Covid-19, including self-isolation. Dividing by the number of NHS employees, 1,169,045, provides a proxy for the probability $p_{12}(t)$ of an NHS employee testing positive. The decline in the fraction of people on Covid-19 sick leave may indicate that NHS England has improved its infection control or it could indicate that more staff have had the disease and generated immunity.

D.3 Finding expressions for the remaining unknowns

It is required to find expressions for the following variables: $n_{11}(t)$, $w_{12}(t)$, $m_{12}(t)$, $m_2(t)$, $p_{11}(t)$, $p_2(t)$.

The situation report gives the number of patients diagnosed on $t_0 = 31$ March 2020, the starting date for this analysis, as $m_{11}(t_0) = 2397$, while the total number of people diagnosed across both Pillars is given by the Coronavirus Dashboard as $m(t_0) = 2414$. Taking the figures at face value, this implies that $m_{12}(t_0) = 17$ people who were not NHS patients, but presumably staff, received tests on 31 March 2020 (and then tested positive). The number of positives cannot exceed the number of tests, of course, so that:

$$m_{12}(t) \le w_{12}(t) \tag{D.9}$$

or

$$m_1(t) - m_{11}(t) \le w_1(t) - w_{11}(t) \tag{D.10}$$

where $m_1(t) = m_{11}(t) + m_{12}(t)$ is the total number of positive cases from Pillar 1 tests on date, t. There was no Pillar 2 testing in the early days and hence, although $m_1(t)$ is not listed explicitly, we may deduce that $m_1(t) = m(t)$ for the first few months. It was then assumed that the number of tests $w_{11}(t)$ given to NHS patients on day t was proportional to the number $n_{adm}(t)$ of patients admitted on that day, provided that this did not imply a total number of tests that is greater than the total for Pillar 1, namely $w_1(t)$. Hence

68 NHS England, COVID-19 Daily Situation Report (12 November 2020) https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-hospital-activity/

$$w_{11}(t)=\begin{cases}\dfrac{w_1(t_0)-\left(m(t_0)-m_{11}(t_0)\right)}{n_{adm}(t_0)}n_{adm}(t) & \text{if } \dfrac{w_1(t_0)-\left(m(t_0)-m_{11}(t_0)\right)}{n_{adm}(t_0)}n_{adm}(t)< \\ & < w_1(t)-\left(m(t)-m_{11}(t)\right) \\ w_1(t)-\left(m(t)-m_{11}(t)\right) & \text{otherwise.}\end{cases} \tag{D.11}$$

The limit is not invoked after 13 April 2020. The number of tests $w_{12}(t)$ applied to NHS England's staff is then found using equation (D.1). Validation for this approach may be found by summing the first 5 days' tests calculated for NHS England workers:

$$\sum_{t=31\text{ Mar }20}^{4\text{ Apr }20} w_{12}(t)=4390, \tag{D.12}$$

which compares well with the stated number: "As of 4th April, nearly 4,500 NHS workers have been tested...", cited above.

Once an estimate is available for the number $w_{11}(t)$ of tests applied to patients, the probability $p_{11}(t)$ of patients testing positive is then:

$$p_{11}(t)=\frac{m_{11}(t)}{w_{11}(t)}. \tag{D.13}$$

A proxy for the probability $p_{12}(t)$ of a positive diagnosis for an NHS staff member may be found as the fraction of NHS England workers on coronavirus sick leave or in isolation.[69] Unsurprisingly,

$$p_{12}(t)<p_{11}(t) \qquad \text{for all } t. \tag{D.14}$$

Given $p_{12}(t)$ and the number of daily tests $w_{12}(t)$, the number $m_{12}(t)$ of NHS England staff being diagnosed each day is given by:

$$m_{12}(t)=p_{12}(t)\,w_{12}(t). \tag{D.15}$$

Now that $m_{12}(t)$ has been established, in addition to $m(t)$ and $m_{11}(t)$, the number of daily positive diagnoses $m_2(t)$ for Pillar 2 may be found from equation (D.7). The probability of a positive result for a test administered in Pillar 2 follows from:

$$p_2(t)=\frac{m_2(t)}{w_2(t)}. \tag{D.16}$$

The total number of positive tests in Pillar 1, $m_1(t)$, is clearly:

$$m_1(t)=m_{11}(t)+m_{12}(t). \tag{D.17}$$

The overall probability of testing positive in Pillar 1 will lie between $p_{12}(t)$ and $p_{11}(t)$. After large scale testing of NHS England staff is introduced, it lies closer to the lower value and may be calculated as:

$$p_1(t)=\max\left(p_{12}(t),\frac{m_1(t)}{w_1(t)}\right). \tag{D.18}$$

[69] NHS England, COVID-19 Daily Situation Report (12 November 2020) https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-hospital-activity/

D.4 Validation of the model against independent data

The modelling approach may be validated against data supplied in the NHS England Test & Trace data tables.[70] Figure D.1 shows the estimated numbers of swab antigen tests processed daily in England, for patients w_{11}, for NHS England workers w_{12}, and for the general public w_2. Figure D.2 shows the number of positive diagnoses, for patients m_{11}, for NHS England workers m_{12}, for NHS patients and workers combined m_1, and for the general public m_2. Also shown are the NHS Test & Trace records, compiled on a weekly basis, of m_1 and m_2. It will be seen that there is good agreement between the model outlined in this Appendix and the Test & Trace (T&T) data. Figure D.3 shows the probability of receiving a positive diagnosis, for patients p_{11}, for NHS England workers p_{12}, for NHS patients and workers combined p_1, and for the general public p_2. Also shown are the NHS Test & Trace records, compiled on a weekly basis, of p_1 and p_2. It will be seen that there is good agreement once again between the model outlined in this Appendix and the Test & Trace data.

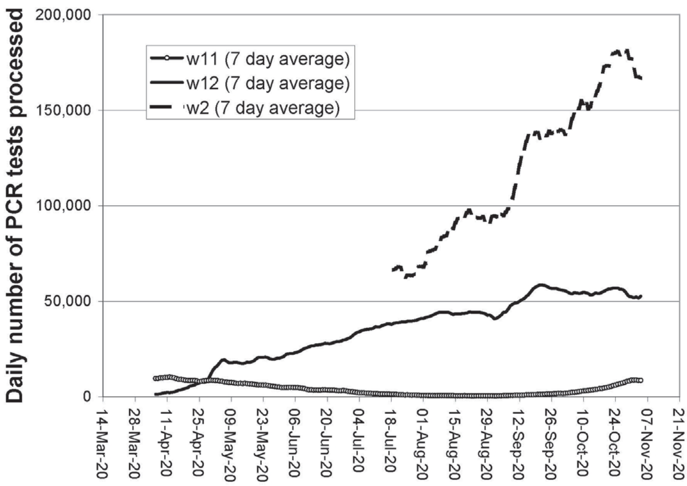

Figure D.1. Daily number of viral tests processed (England). w_{11} is the estimated number of tests on hospital patients, w_{12} is the estimated number of tests on NHS England staff and w_2 is the reported number of tests on symptomatic members of the general public (Pillar 2).

D.5 Use of the model and the Test & Trace data

The two methods of estimation, modelling and the Test and Trace records produce comparable results. The basic requirement of the measuring system can mainly be satisfied by using the data on p_1 and p_2 supplied by Test and Trace, but extended into earlier months using the model presented here. The behaviour of p_2 during April 2020, based on the model, receives support from the findings of an independent study analysing data from 2.5 million people in the UK

[70] NHS T&T data tables (week 23, 2020) https://www.gov.uk/government/publications/nhs-test-and-trace-england-and-coronavirus-testing-uk-statistics-29-october-to-4-november/weekly-statistics-for-nhs-test-and-trace-england-and-coronavirus-testing-uk-29-october-to-4-november

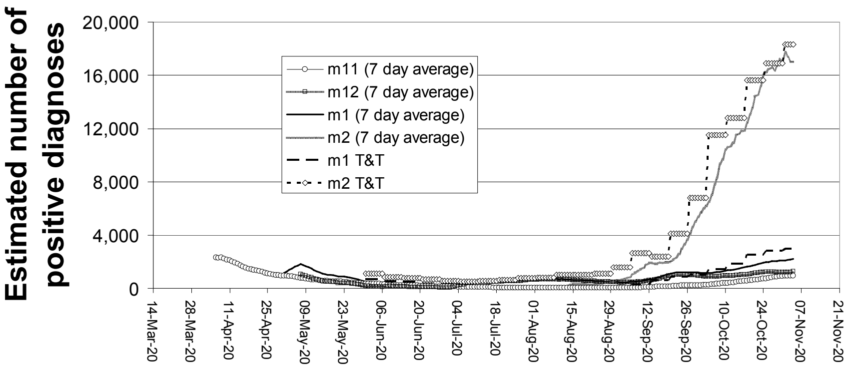

Figure D.2. The number of new cases daily, calculated versus figures available from NHS Test & Trace. m_{11} is the estimated number of new diagnoses in hospital patients, m_{12} is the estimated number of new diagnoses in NHS England staff, m_1 is the sum, $m_1 + m_2$, m_2 is the estimated number of new diagnoses in asymptomatic members of the public (Pillar 2), m_1 T&T is the total number of NHS England patients and staff testing positive per day reported by NHS Test & Trace and m_2 T&T is the number of symptomatic members of the public testing positive.

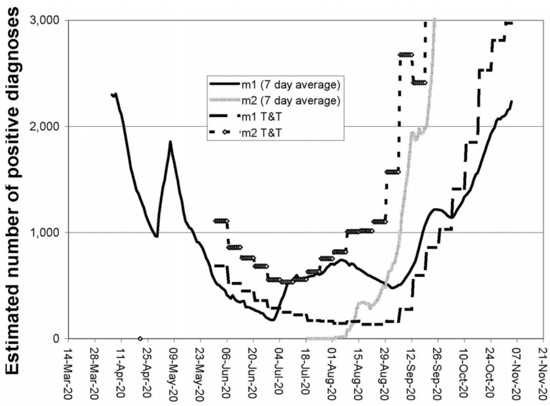

Figure D.3. The number of new cases daily, calculated versus figures available from NHS Test & Trace (Expanded scale). m_1 is the total number of NHS patients and staff testing positive per day (Pillar 1), m_2 is the estimated number of new diagnoses in asymptomatic members of the public (Pillar 2), m_1 T&T is the total number of NHS England patients and staff testing positive per day reported by NHS Test & Trace and m_2 T&T is the number of symptomatic members of the public testing positive.

reporting respiratory symptoms between 24 March and 21 April 2020.[71] That research found that 17.42% of people with such symptoms would have Covid-19. This figure is close to the average found from the model from 31 March to 21 April 2020, namely 15.5%.

[71] Menni, C. et al., Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nature Medicine* 26 (2020) 1037–1040 (https://www.nature.com/articles/s41591-020-0916-2).

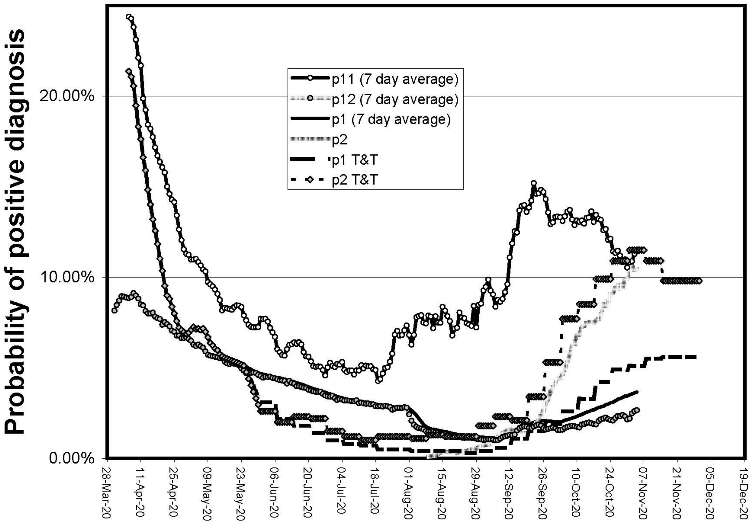

Figure D.4 Model probabilities compared with those from Test & Trace (T&T). Given that a person is being tested, p_{11} is the probability of an NHS patient testing positive, p_{12} is the probability of an NHS staff member testing positive, p_1 is the combined probability of an NHS patient or staff member testing positive, while p_2 is the probability of a member of the public testing positive.

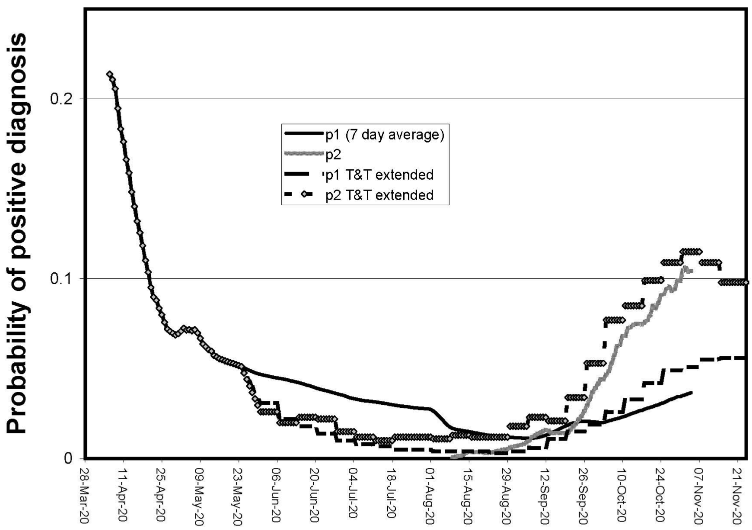

Figure D.5. Probabilities of testing positive in Pillar 1, p_1, and Pillar 2, p_2, given that the person is tested, where Pillar 1 applies to NHS staff and patients, while Pillar 2 applies to the general public. Model values for the probability p_1 of an NHS patient or staff member testing positive under test and for the probability p_2 of a member of the public testing positive under test are available for the period from late March onwards. These have been used to extend the graphs of p_1 and p_2 based on Test & Trace (T&T) figures to cover the early stages of the epidemic, where T&T figures are not available.

Appendix E. Expected value and variance of the estimate of the social distancing index, R_0, found from ONS surveys

The test for T-cell immunity was based on a sample, $n = 2{,}197$, of seronegative people, of whom $x = 283$, or 12.9%, were found to have T-cell immunity. The fraction of people with T-cell immunity may thus be regarded a binomial parameter, $p_T = P$ (T-cell immunity), where

$$E(P_T) = \frac{x}{n} = 0.129 \tag{E.1}$$

and

$$\mathrm{var}(P_T) = \sigma_{PT}^2 \approx \frac{\frac{x}{n}\left(1 - \frac{x}{n}\right)}{n} = 5.1079 \times 10^{-5}. \tag{E.2}$$

The ONS finds the fraction with antibodies, P_{Ai}, at roughly monthly intervals for specified dates. It is assumed that this slowly changing value is valid in the interval between dates, i and $i+1$, so that

$$P_A(t) = P_{Ii} \quad \text{for } i \le t < i+1. \tag{E.3}$$

$P_A(t)$ is a random variable with expected value $E(P_A(t))$ and variance $\sigma_{PA}^2(t)$ for dates t as shown in Table E.1. The fraction $P_I(t)$ of people who are immune on the various dates is the sum of those with T-cell immunity and those with antibody immunity:

$$P_I(t) = P_T + P_A(t). \tag{E.4}$$

$P_I(t)$ is a random variable, for which the expectation is:

$$E(P_I(t)) = \mu_{PI} = E(P_T) + E(P_A(t)). \tag{E.5}$$

Table E.1. ONS survey estimates of the fraction of people with Covid-19 antibodies.

Date (t)	95% confidence interval, lower level	Central value $E(P_A(t))$	95% confidence interval, upper level	Variance $\sigma_{PA}^2(t)$
29 June 2020	0.050	0.063	0.078	5.1020×10^{-5}
26 July 2020	0.050	0.062	0.076	4.3992×10^{-5}
08 Sep. 2020	0.054	0.062	0.071	2.2053×10^{-5}
30 Sep. 2020	0.050	0.056	0.062	9.3711×10^{-6}

Meanwhile the variance is simply the sum of the variances, because the probabilities are independent:

$$\sigma_{PI}^2(t) = \sigma_{PT}^2 + \sigma_{PA}^2(t). \tag{E.6}$$

The social distancing index, $R_0(t)$, derived from ONS data may be computed as:

$$R_0(t) = \frac{R(t)}{1 - P_I(t)} \tag{E.7}$$

where $R(t)$ and $R_0(t)$ are random variables, as measured by survey.

As noted in Appendix C, if $Y = h(X_1,X_2)$ is a nonlinear function of two random variables X_1 and X_2 with means μ_{X1} and μ_{X2} and variances σ^2_{X1} and σ^2_{X2}, then

$$E(Y) \approx h(\mu_{X1},\mu_{X2}) + \frac{1}{2}\left(\frac{\partial^2 h(\mu_{X1},\mu_{X2})}{\partial X_1^2}\sigma^2_{X1} + 2\frac{\partial^2 h(\mu_{X1},\mu_{X2})}{\partial X_1 \partial X_2}\sigma_{X1,X2} + \frac{\partial^2 h(\mu_{X1},\mu_{X2})}{\partial X_2^2}\sigma^2_{X2}\right) \quad \text{(C.3)}$$

where $\sigma_{X1,X2} = E\left((X_1 - \mu_{X1})(X_2 - \mu_{X2})\right)$ is the covariance between X_1 and X_2. For the case where $h(X_1,X_2) = X_2/X_1$, and $X_1 = 1 - P_I(t)$

$$\begin{aligned}
&\frac{\partial h}{\partial X_1} = \frac{-X_2}{X_1^2}; \quad \frac{\partial h}{\partial X_2} = \frac{1}{X_1} \\
&\frac{\mathrm{d}X_1}{\mathrm{d}P_I} = -1; \quad \frac{\partial h}{\partial P_I} = \frac{X_2}{(1-P_I)^2} \\
&\frac{\partial^2 h}{\partial P_I^2} = 2\frac{X_2}{(1-P_I)^3}; \quad \frac{\partial^2 h}{\partial X_2^2} = 0; \quad \frac{\partial h}{\partial P_I \partial X_2} = \frac{1}{(1-P_I)^2}.
\end{aligned} \quad \text{(E.8)}$$

In this case, where $X_1 = R(t)$ there is only a weak theoretical connection between $P_I(t)$ and $R(t)$, which may be confirmed empirically by plotting the two against each other. It will be assumed that the covariance is zero. Hence

$$E(R_0(t)) = E\left(\frac{R(t)}{1-P_I(t)}\right) \approx \frac{\mu_R}{1-\mu_{PI}} + \frac{\mu_R}{(1-\mu_{PI})^3}\sigma^2_{PI} \approx \frac{\mu_R}{1-\mu_{PI}}\left(1 + \frac{\sigma^2_{PI}}{(1-\mu_{PI})^2}\right) \quad \text{(E.9)}$$

where $\mu_R = E(R(t))$. As noted in Appendix C, the variance for the nonlinear function $Y = h(X_1,X_2)$, may be approximated by:

$$\begin{aligned}
\mathrm{var}(Y) \approx \left(\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_1}\right)^2 \sigma^2_{X1} + 2\rho_{X1,X2}\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_1}\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_2}\sigma_{X1}\sigma_{X2} + \\
\left(\frac{\partial h(\mu_{X1},\mu_{X2})}{\partial X_2}\right)^2 \sigma^2_{X2}
\end{aligned} \quad \text{(C.11)}$$

where ρ_{X_1,X_2} is the correlation coefficient between X_1 and X_2. In the current case, where the correlation coefficient is assumed to be zero:

$$\mathrm{var}(R_0(t)) \approx \left(\frac{\partial h}{\partial P_I}\right)^2 \sigma^2_{PI} + \left(\frac{\partial h}{\partial R}\right)^2 \sigma^2_R \approx \left(\frac{\mu_R}{1-\mu_{PI}}\right)^2 \frac{\sigma^2_{PI}}{(1-\mu_{PI})^2} + \frac{\sigma^2_R}{(1-\mu_{PI})^2}. \quad \text{(E.10)}$$

Appendix F. Allowing for vaccination

The government's policy is to assume that everyone is equally vulnerable to infection, whether or not they have contracted or recovered from Covid-19. This was illustrated by the Prime Minister going into self-isolation on 16 November 2020, having come into contact on 12 November 2020 with an MP who subsequently tested positive, and emerging from quarantine on 26 November 2020.[72,73] While the caution shown in that instance might be regarded as excessive and against the evidence to date—there have been only a handful of cases, out of the many millions recorded worldwide, where a repeat infection occurred—it may be a sensible policy stance to adopt for a vaccination programme. While millions of people have been infected and recovered in the UK, many of these will have no sure proof or record that it was indeed Covid-19 from which they suffered, and the majority will not know that they were infected in the past; nor will those with T-cell immunity know that they possess this characteristic. Hence the Government will have little choice but to offer the vaccine to everyone, irrespective of whether they are already immune or not. A substantial fraction of vaccine doses will be "wasted" as a result, in the sense of having no beneficial effect.

The assumption will be made that England will begin vaccinating people on t_{vac0} = 14 December 2020, but that there will be a delay of τ_{vac} days before immunity can be attained. The effectiveness of the vaccination in conferring immunity will be η_{vac}. It will be assumed that vaccinations occur at a daily rate of $v(t)$, so that the total number of vaccinations at any time, $n_{vaccT}(t)$, will be:

$$n_{vaccT}(t) = \int_{t_{vac0}}^{t} v(t)\mathrm{d}t \tag{F.1}$$

It will be assumed that the vaccination will be given without discrimination to all except those with a clear active infection, taken to imply Cohort 1. It will be further assumed that the people who have been vaccinated may be regarded as a subset of those categorized as recovered. Hence the number of vaccinations $v_{si}(t)$ given each day to those in Cohort i who are susceptible will be

$$v_{si}(t) = \frac{n_{si}(t)}{N - n_1(t) - n_{Tvac}(t)} v(t) \tag{F.2}$$

where $n_{si}(t)$ is the number of people in Cohort i who are still susceptible, $n_i(t)$ is the number of people in Cohort i with an active infection and N is the total number of people in the population. The number of (susceptible) people becoming immune in each cohort from vaccination each day, $q_i(t)$, will be

[72] Roach, A., Boris Johnson pictured outside Downing Street as self-isolation comes to an end. *Evening Standard* (26 November 2020) https://www.standard.co.uk/news/politics/boris-johnson-out-of-isolation-pictures-b79282.html

[73] The Prime Minister's self-isolation despite having contracted and recovered from Covid-19 earlier in the year would seem to go against the basic assumption of all epidemiological modelling of Covid-19 to date; one notes, however, that there have been a handful of cases, out of the many millions recorded worldwide, where a repeat infection has occurred.

$$q_i(t) = \begin{cases} 0 & \text{for } t < t_0 + \tau_{vac} \\ \eta_{vac} v_{si}(t - \tau_{vac}) & \text{for } t \geq t_0 + \tau_{vac}. \end{cases} \quad \text{(F.3)}$$

The new rate of increase of those classified as recovered is now augmented by adding $q_i(t)$ to the right hand side of equation (A.6) of Thomas[74]

$$\frac{\mathrm{d}n_{r,i}}{\mathrm{d}t} = \frac{n_i(t)}{\tau_{\inf,i}} + q_i(t). \quad \text{(F.4)}$$

Meanwhile equation (A.7) of the same paper needs to be replaced by:

$$\frac{\mathrm{d}n_i}{\mathrm{d}t} = \frac{\mathrm{d}n_{x,i}}{\mathrm{d}t} - \frac{n_i(t)}{\tau_{\inf,i}} \quad \text{(F.5)}$$

as, by equation (F.4) above, $\mathrm{d}n_{r,i}/\mathrm{d}t$ is no longer equal to $n_i(t)/\tau_{\inf,i}$.

Appendix G. Allowing for the vaccination of the vulnerable first

The Secretary of State for Health, Matt Hancock, stated in an interview reported on 6 December 2020 that more than half of vulnerable people will be vaccinated by the end of February 2021.[75] Based on ONS life tables, $f_A = 79\%$ of the population of England will be 65 or under, implying that the complementary fraction, $f_B = 21\%$, will be over 65. People over 65 account for 89.3% of the deaths from Covid-19, as explained in §2.1. In general, the infection fatality rate, I, will be given as:

$$I = \frac{n_{dA} + n_{dB}}{n_A + n_B} = \frac{n_{dA}}{n_A} \frac{1 + \frac{n_{dB}}{n_{dA}}}{1 + \frac{n_B}{n_A}} \quad \text{(G.1)}$$

where n_A is the number of those infected and recovered[76] aged 65 or under, n_B is the number of those infected and recovered who are over 65, n_{dA} is the number of people aged 65 or under who have died and n_{dB} is the number of those who aged over 65 who have died. n_A includes those in both Cohorts 1 and 2 as long as they are 65 or below; similarly n_B includes those in both Cohorts 1 and 2 who are aged over 65. If it is assumed, for simplicity, that infections are spread equally across all ages, then

$$\frac{n_A}{n_B} = \frac{f_{AR} n}{f_{BR} n} = \frac{f_A}{f_B} \quad \text{(G.2)}$$

---

74 Thomas, P., 2020, "The length and severity of the coronavirus recession estimated from the dynamics of relaxing lockdown", *Nanotechnology Perceptions*, http://www.colbas.org/ntp/opnAxs/N07TH20A.pdf

75 Hope, C., Hancock: vaccine will loosen tiers before end of March. *Sunday Telegraph* (6 December 2020) https://digitaleditions.telegraph.co.uk/data/438/reader/reader.html?#!preferred/0/package/438/pub/438/page/3/article/110711

76 "Recovered" is taken to include those who have unfortunately succumbed to the disease.

where f_{AR} is the fraction aged 65 or under among the total number $n = n_A + n_B$ of people who had been infected and recovered, while f_{BR} is the corresponding number aged over 65. Meanwhile, the ratio of deaths may be expressed as

$$\frac{n_{dB}}{n_{dA}} = \frac{f_{dB} n_d}{f_{dA} n_d} = \frac{f_{dB}}{f_{dA}} \tag{G.3}$$

where n_d is the number who have died, f_{dA} is the fraction of the dead who were 65 or below and f_{dB} is the fraction of the dead who were over 65. Substituting into equation (G.1) and rearranging then gives the infection fatality rate, I_A, for those aged 65 or under:

$$I_A = \frac{n_{dA}}{n_A} = \frac{1 + \dfrac{f_B}{f_A}}{1 + \dfrac{f_{dB}}{f_{dA}}} I = \frac{\dfrac{f_A + f_B}{f_A}}{\dfrac{f_{dA} + f_{dB}}{f_{dA}}} I = \frac{f_{dA}}{f_A} I \tag{G.4}$$

where equation (G.2) has been used and the identities, $f_A + f_B = 1 = f_{dA} + f_{dB}$, have been employed in the last step. Analogously, the infection fatality rate I_B for those over 65 is

$$I_B = \frac{f_{dB}}{f_B} I. \tag{G.5}$$

Before vaccination, time t_{vac0}, $f_A = f_A(t_{vac0}) = 0.79$ and $f_B = f_B(t_{vac0}) = 0.21$.

Moreover, 89.3% of the Covid-19 deaths occur in the over 65s and so 10.7% occur in those aged 65 or under. Taking the infection fatality rate for the unvaccinated population as $I_0 = 0.005$, the infection fatality rate for those aged up to and including 65 is

$$I_A = \frac{0.107}{0.79} \times 0.005 = 6.77 \times 10^{-4} \tag{G.6}$$

while the infection fatality rate for those aged over 65 is

$$I_B = \frac{0.893}{0.21} \times 0.005 = 2.13 \times 10^{-2}. \tag{G.7}$$

The figures for I_A and I_B may be regarded as invariant over time. However, the overall infection fatality rate, I, for a population made with a different split between ages will vary depending on the values of f_A and f_B. In particular, f_B will be a decreasing number if the vaccination is given preferentially to those over 65.

Equation (G.1) may be rewritten:

$$I = \frac{n_{dA} + n_{dB}}{n_A + n_B} = \frac{n_{dA}}{n_A} \frac{n_A}{n_A + n_B} + \frac{n_{dB}}{n_B} \frac{n_B}{n_A + n_B} \tag{G.8}$$

or

$$I(t) = f_A(t) I_A + f_B(t) I_B \tag{G.9}$$

or, noting that $f_A(t) + f_B(t) = 1$,

$$I(t) = I_A + f_B(t)(I_B - I_A). \tag{G.10}$$

There are $N_B\left(t_{vac0}\right) = 0.21 \times 56.5$ million = 11.87 million people in the population of England who are over 65 at the start of the vaccination exercise. In Scenarios 1 and 2 is assumed that the first 10 million vaccinations are given to people in this age group, implying a vaccination percentage of $10\times10^6/11.87\times10^6 = 84\%$. 89.3% of the deaths come from this age group, and so it can be assumed that 75% of those at risk of dying will have been vaccinated.

At the assumed vaccination rate of 5 million per month, these people will have received their first vaccination by the end of February 2021. Completion of vaccination is assumed to be delayed by $\tau_{vac} = 28$ days on the first vaccination. Hence the rate of reduction of the effective number of people in the population above the age of 65 for any date, t, for the first three months from the start vaccination is given by $v(t-\tau_{vac})$, where $v(t)$ is the number of 1st vaccinations per day (see Appendix F).

The number of people aged 65 or above in England whose vaccination is complete will therefore be

$$\Delta N_B\left(t\right) = \int_{t=t_{vac0}}^{t} v\left(t-\tau_{vac}\right)\mathrm{d}t \qquad \text{for } t_{vac0} \le t \le t_{Bfin} \tag{G.11}$$

So that the number of people over 65 at any time, t, whose vaccinations are incomplete will be

$$N_B\left(t\right) = N_B\left(t_{vac0}\right) - \Delta N_B\left(t\right) \tag{G.12}$$

The fraction of people who are over 65 and whose vaccination is incomplete will thus be

$$f_B\left(t\right) = \frac{N_B\left(t\right)}{N_A\left(t\right) + N_B\left(t_{vac0}\right)} \qquad \text{for } t_{vac0} < t \le t_{Bfin} \tag{G.13}$$

where t_{Bfin} marks the end date of preferential vaccination for people aged 65 and above plus 28 days. This is assumed to be t_{Bfin} = 25 March 2021. It is assumed that the first vaccination ceases to be administered preferentially on 25 February 21, after 84% of the cohort aged 65 and over have received an injection. Hence

$$f_B\left(t\right) = \frac{N_B\left(t_{Bfin}\right)}{N_A\left(t_{Bfin}\right) + N_B\left(t_{vac0}\right)} \qquad \text{for } t > t_{Bfin}. \tag{G.14}$$

The value of the infection fatality rate on date, t, is found by combining (G.10), (G.13) and (G.14).

The rate of people dying on date, t, is then the number of people recovering on that date multiplied by the current infection fatality rate:

$$\frac{\mathrm{d}n_d}{\mathrm{d}t}\left(t\right) = I\left(t\right)\left(\frac{n_1\left(t\right)}{\tau_{\inf,1}} + \frac{n_2\left(t\right)}{\tau_{\inf,2}}\right). \tag{G.15}$$