Estimating R for the UK using Publicly Available Data

By Stuart McDonald for

COVID-19 Actuaries Response Group – Learn. Share. Educate. Influence.

Summary

Fundamental to the spread of any virus is the reproductive rate ("R"). When this is above 1.0 then each infected person infects more than one other person, and the infectious population grows; below 1.0 and it falls.

As countries around the world begin to relax social distancing policies it is interesting to estimate R using publicly available data. Using such estimates we can tentatively address questions such as what might have happened if certain measures had been introduced earlier, later or not at all – though we save detailed consideration for a subsequent bulletin.

The most reliable publicly available data for this purpose are time series of deaths. By assessing the rate of change in deaths we can estimate the rate of change in infections at an earlier date. Unfortunately, this approach means that our estimated R value cannot be completely up to date.

Our approach appears reliable when R is stable. However, it is not able to reliably identify dates at which significant changes occurred. This is a primarily a consequence of significant variation in the time between individuals becoming infected and dying. Any significant actual fall in R would show up in our historical analysis as a gradual fall starting around a week earlier. It is likely that other estimates of R over March may suffer from similar limitations, casting doubt on the validity of statements that R was starting to drop below 1 by 23 March.

We will continue to update our estimates so that we can observe how R is changing over time.

The Rate of Reproduction

Key to understanding the spread of a virus is the reproductive rate "R". This rate varies over time so we refer to the rate a time t as R_t (R_0 is the R value at the start of the epidemic).

To measure R_t in real time requires extremely widespread testing at a scale which has not been available in the UK. Instead we can look at the rate at which events such as hospital or intensive care admissions and deaths vary over time.

R is of course a simplification and there are clearly differences in the distribution of the virus by geography, age, sex and other factors. Nonetheless, we believe our approach to estimating R gives us a reasonable picture of how the rate of transmission is varying over time in the UK as a whole. We will provide more detail of our method of R estimation in a future paper, including details of how actuarial techniques can be used to produce a more up-to-date estimate than is used in this bulletin.

Data Sources

Useful sources of reliable data are:

- Deaths of patients in hospitals in England who had tested positive for COVID-19 or where COVID-19 was mentioned on the death certificate produced by NHS England
- Deaths where COVID-19 was mentioned on the death certificate for England and Wales produced by the Office for National Statistics (ONS)

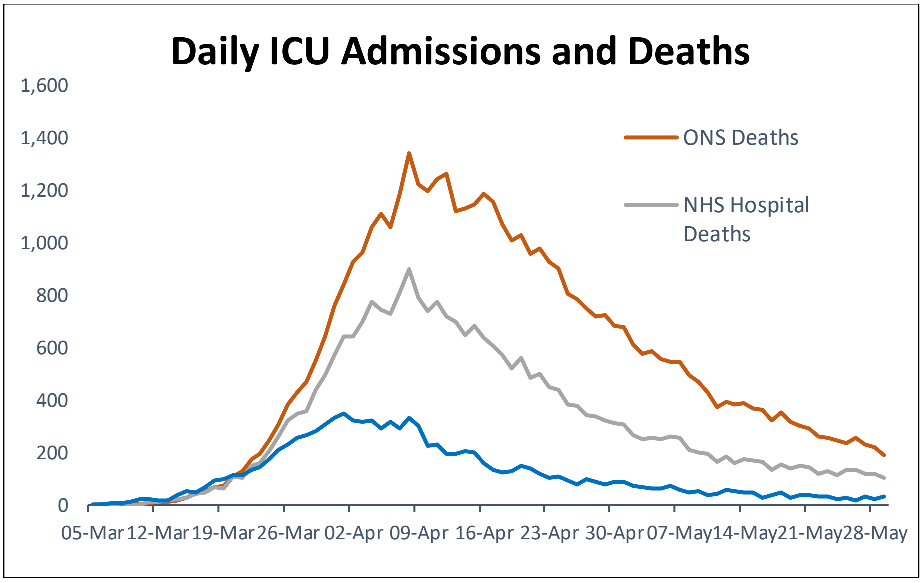
In this bulletin we use data from:

- NHS deaths registered by 8 June (published 9 June)
- ONS weekly deaths registered by 29 May (published 9 June)

These reports contain daily time series which can be used to examine how deaths are varying by day. By choosing the ONS reporting by date of death rather than date of notification we avoid distortions caused by weekends and public holidays. NHS reporting is also by date of death.

Deaths peaked on 8 April with 1,341 deaths total deaths in England and Wales where COVID-19 was mentioned on the death certificate (ONS). 899 of these were in English hospitals (NHS).

We looked also at ICU admissions, applying similar techniques to see if we could get more up to date R_t estimates. The estimates were consistent when daily admissions were high but became noisy as daily admissions fell through April.



Estimating R

We can estimate the reproductive rate of the virus at a historical point by examining how the number of deaths was changing, after allowing for the average interval between infection and death. We assume a 19-day interval between infection and death, consistent with the 26th Report from Imperial College London (ICL).

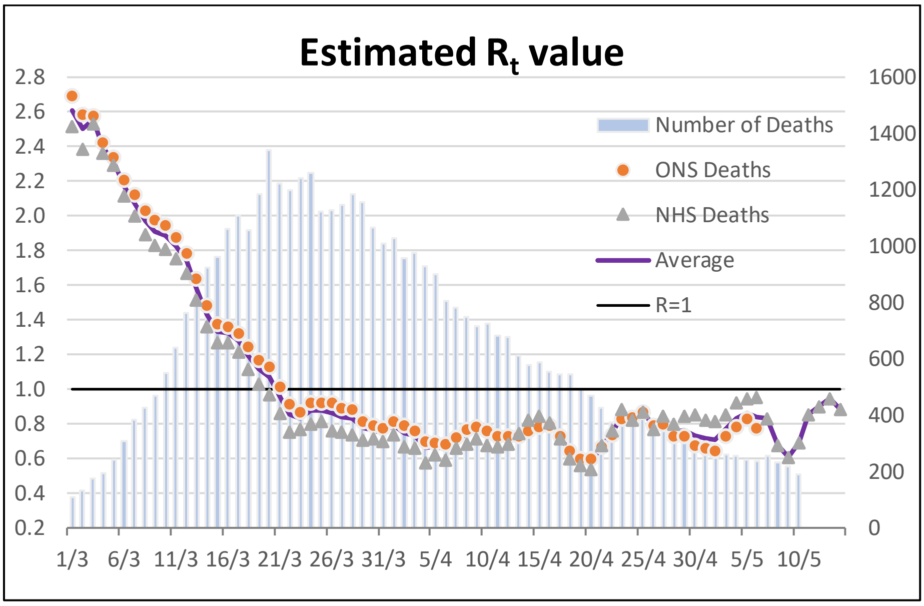
R is a function of the ratio between the number of new deaths and the equivalent number at an earlier point. In order to reduce the volatility of our estimate caused by random fluctuations we use a measurement interval of five days and average over a three-day period.

Estimated R_t values are illustrated in the following chart, showing values estimated from ONS and NHS data and a weighted average of the two estimates.

Observations

Several observations can be made from our estimated R_t values.

The estimates are reassuringly consistent with R_t clearly falling over the period in question. It was around 2.5 at the start of March and was in the range 0.6 to 0.8 throughout April.



It appears that R may be rising back towards 1.0 at the end of the period for which we can currently estimate it. This is the weekend in which the VE Day Bank Holiday fell and the emphasis from Government relaxed from “stay at home” to “stay alert”.

There are no clear step changes in the data as might initially be expected in response to the introduction of control measures. There are several reasons for this including:

- Natural variation in the gap between becoming infected and infecting others
- Natural variation in the gap between becoming infected and being admitted to ICU or dying
- Control measures were introduced in stages, often just a couple of days apart
- Different groups responded more or less rapidly to control measures; for example, large employers were moving towards working from home before the instruction to “stay at home” was given by Government
- Our averaging approach smooths the underlying time series
- The transmission rate naturally falls as the susceptible population falls. This has only a very small effect over the period considered.

The relationship between the R estimates is interesting. We know that the average age at death in hospitals is lower than for deaths occurring elsewhere, for example in care homes. We can therefore surmise that when the estimate of R from NHS data exceeds that from ONS data, as in the latter period, then R is higher among the working age population. When the estimate from ONS data is higher then R is higher among the older and more frail.

Timeline of events likely to have influenced R

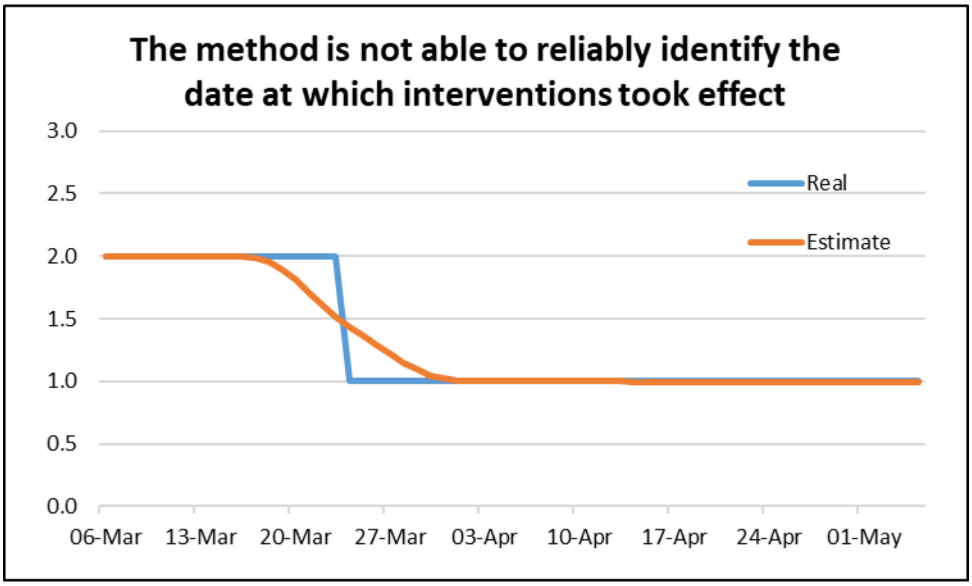
31 January	First two UK cases. First evacuation flight.
28 February	First known community transmission case in the UK.
4 March	First UK deaths.
12 March	People with a cough or fever told to self-isolate.
16 March	Non-essential travel/contact to be avoided.
18 March	Announced that schools will close that week.
20 March	Schools/pubs/restaurants close.
23 March	Instruction to “stay at home”.
10 May	Change in emphasis to “stay alert”.

Limitations

As noted above, this method results in apparently gradual changes in R over time. This means that it is not possible to reliably identify the date at which interventions have had an effect. To demonstrate this we have created an artificial projection of infections and deaths, setting R to be 2.0 until 23 March and 1.0 thereafter. We have then applied our method to estimate R to see how close we get to the actual answer (the actual answer in this artificial exercise being "2.0 becomes 1.0 on 24 March").

As can be seen on the graph, the estimate produces a smooth transition beginning before the actual fall in R. This could lead to an incorrect conclusion that interventions were unnecessary as R had already fallen before they were introduced.

It is likely that other estimates of R over March may suffer from similar methodological limitations, casting doubt on the validity of statements that R was starting to drop below 1 by 23 March.



Next Steps

Having estimated R at different points in time, we could now produce scenarios estimating how different the number of infections (and ultimately deaths) would be if R had stopped falling.

We can apply similar logic if R_t begins to rise as social distancing measures are relaxed, having been clearly under control for nearly two months. We can also produce regional estimates using the breakdown of NHS deaths (noting that smaller numbers mean these will be less certain).

It is clear that infections and deaths are highly sensitive to relatively small changes in R. Great care will be required, and close monitoring of early warning indicators of change, as we relax control measures. We will continue to update and share our estimates of Rt over the weeks ahead.

References

https://www.icnarc.org/Our-Audit/Audits/Cmp/Reports

https://www.england.nhs.uk/statistics/statistical-work-areas/covid-19-daily-deaths/

https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/datasets/weeklyprovisionalfiguresondeathsregisteredinenglandandwales

https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-06-08-COVID19-Report-26.pdf