EXCELling with Covid-19 data

David Wallace

(djw75@cam.ac.uk; https://www.chu.cam.ac.uk/people/view/david-wallace/)

Disclaimers:

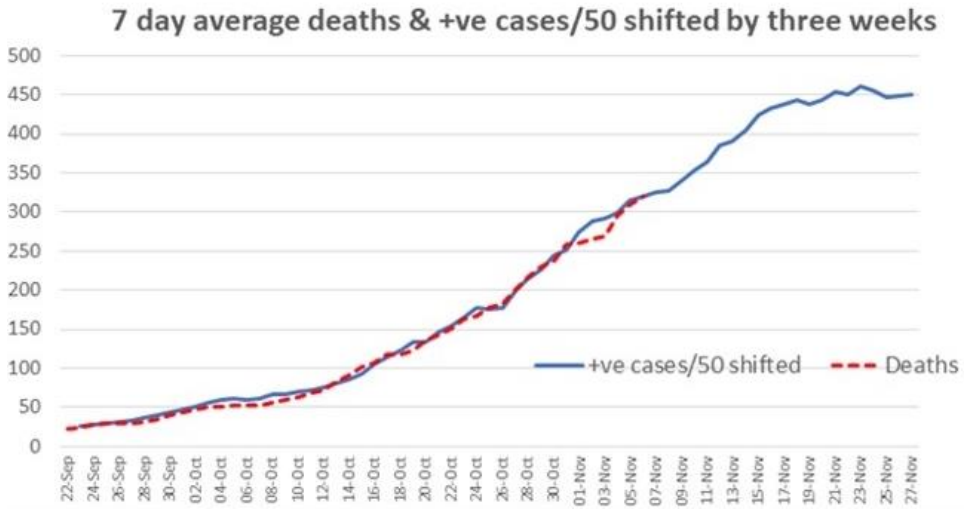
- **Title is a pun, not a boast**
- **I am a volunteer in RAMP, thanks to Graeme Ackland**

Outline:

1. **How I got involved**
2. **Case Fatality Rates from public data**
3. **R-value and growth rate**
4. **Euro2020**

1. How I got involved

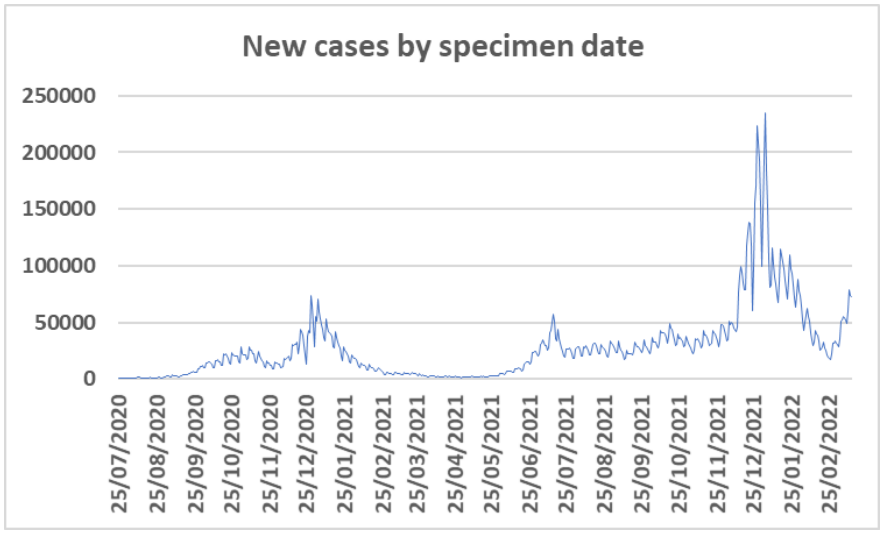
- Vallance and Whitty chart 31 Oct 2020 20201030 GCSA slides for 31 Oct presser FINAL slides used - corrected - Read-Only (1) (publishing.service.gov.uk) Scenarios of 1000-6000 deaths per day by end Dec. (Note, all pre-Alpha…)
- David Spiegelhalter Spectator TV interview, 5 Nov 2020: 'rule of thumb 1 in 50 cases result in death in 3-4 weeks'.
- Email to DS, 8 Nov: deaths peak at c. 450 per day, by end Nov? Actual peak was 466 on 22 Nov (pre-Alpha)
- HEALTH WARNING! The more I know, the less confident I am about making predictions.

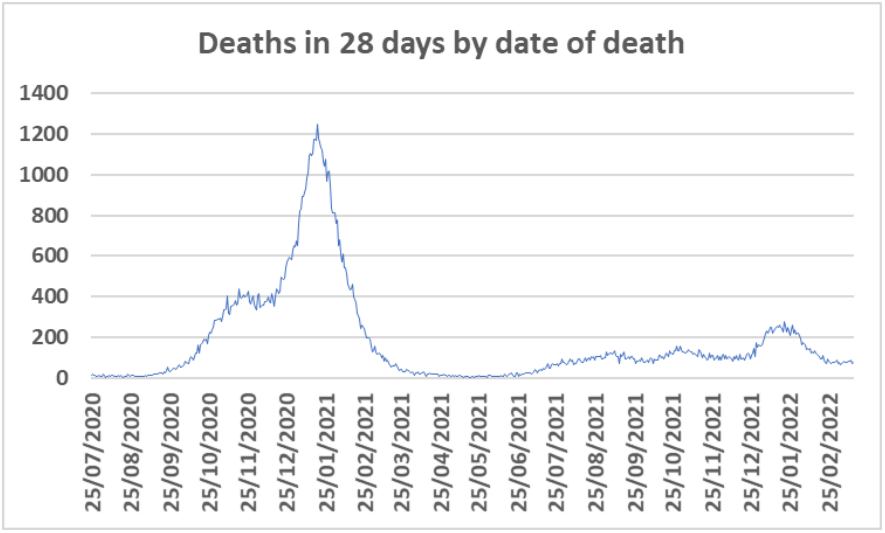


- Absurdly simplified: must add distribution for time to death ('**W**eight and **S**hift'), age dependence for likelihood of death ('**S**cale') etc etc.

2. **Case Fatality Rates from public data**

- Some positive tests result in death.





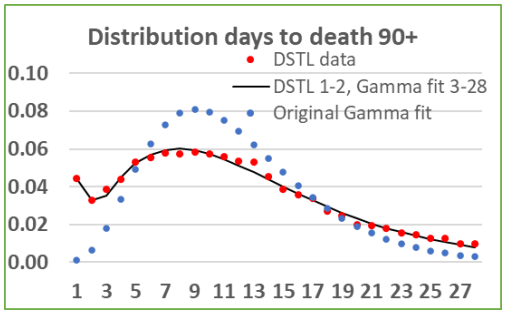
https://coronavirus.data.gov.uk/details/download

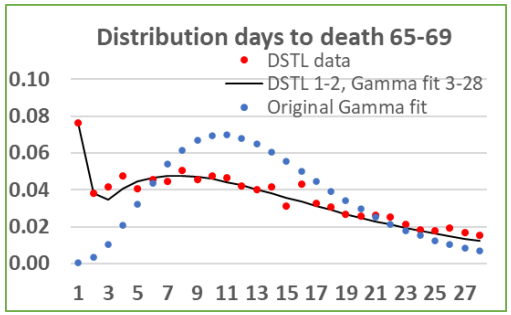
- Aim: map cases onto deaths through Case Fatality Rate (CFR)
- Copy of Excel workbook control panel shows parameters used:

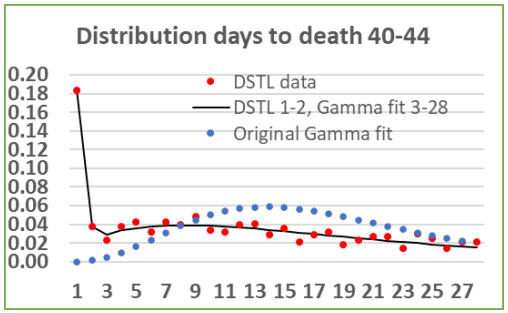
False positives =
B VOC =
Alpha VOC =
Delta VOC =
AY.4.2 VOC =
Omicron VOC =
Reduction first vacc. =
Reduction second vacc. =
Reduction third vacc. =

2. **Case Fatality Rates from public data cont.**

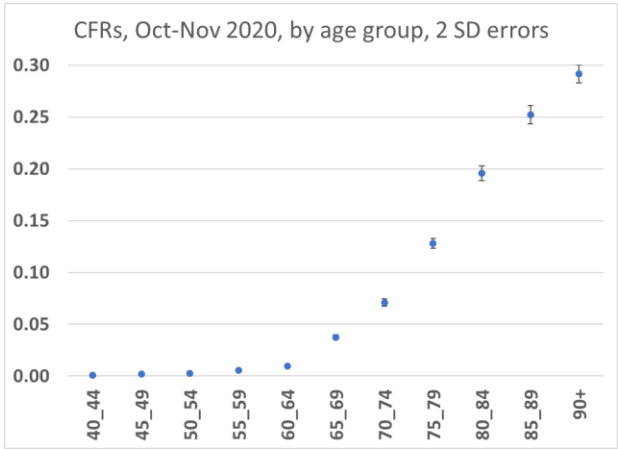
- Cases are by specimen date; deaths are within 28 days by date of death.
- Must build in distribution of time to death ('**W**eight and **S**hift'). Initially we fitted with lognormal and gamma distributions (indistinguishable…). Now based on (confidential) DSTL data (via GJA). Use Gamma distribution, apart from days 1 and 2 *(tested positive on admission for other reason – died 'with' not 'of' Covid-19?) :*





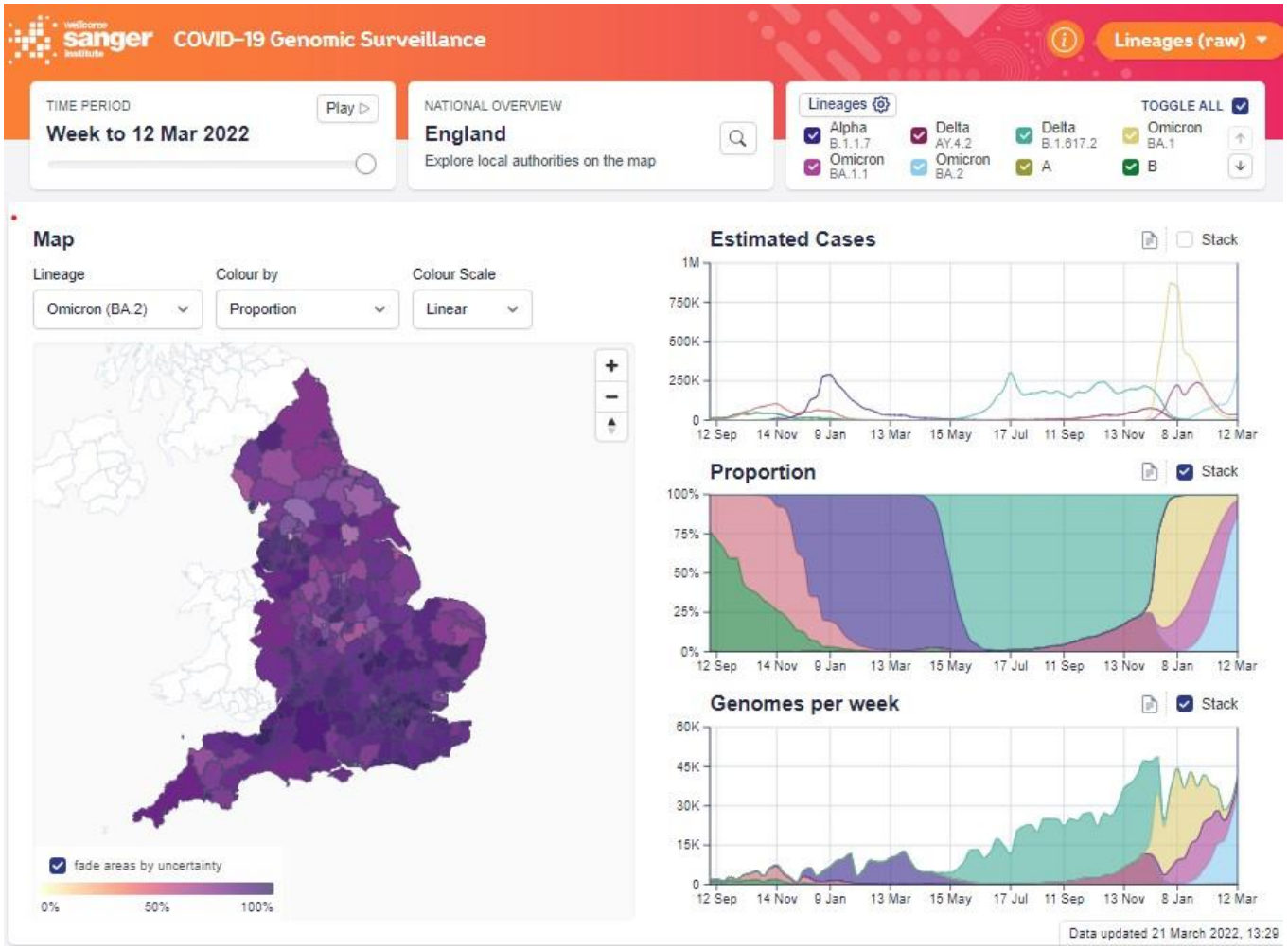


- The CFRs are the **S**cale factors in **WSS**. Estimated by least squares minimising (Actual *minus* Cases-expected deaths), or LINEST function in Excel
- Use Oct-Nov 2020 data (B.1.177) as base-line.
- CFRs are very strongly age dependent.



2. **Case Fatality Rates from public data cont.**

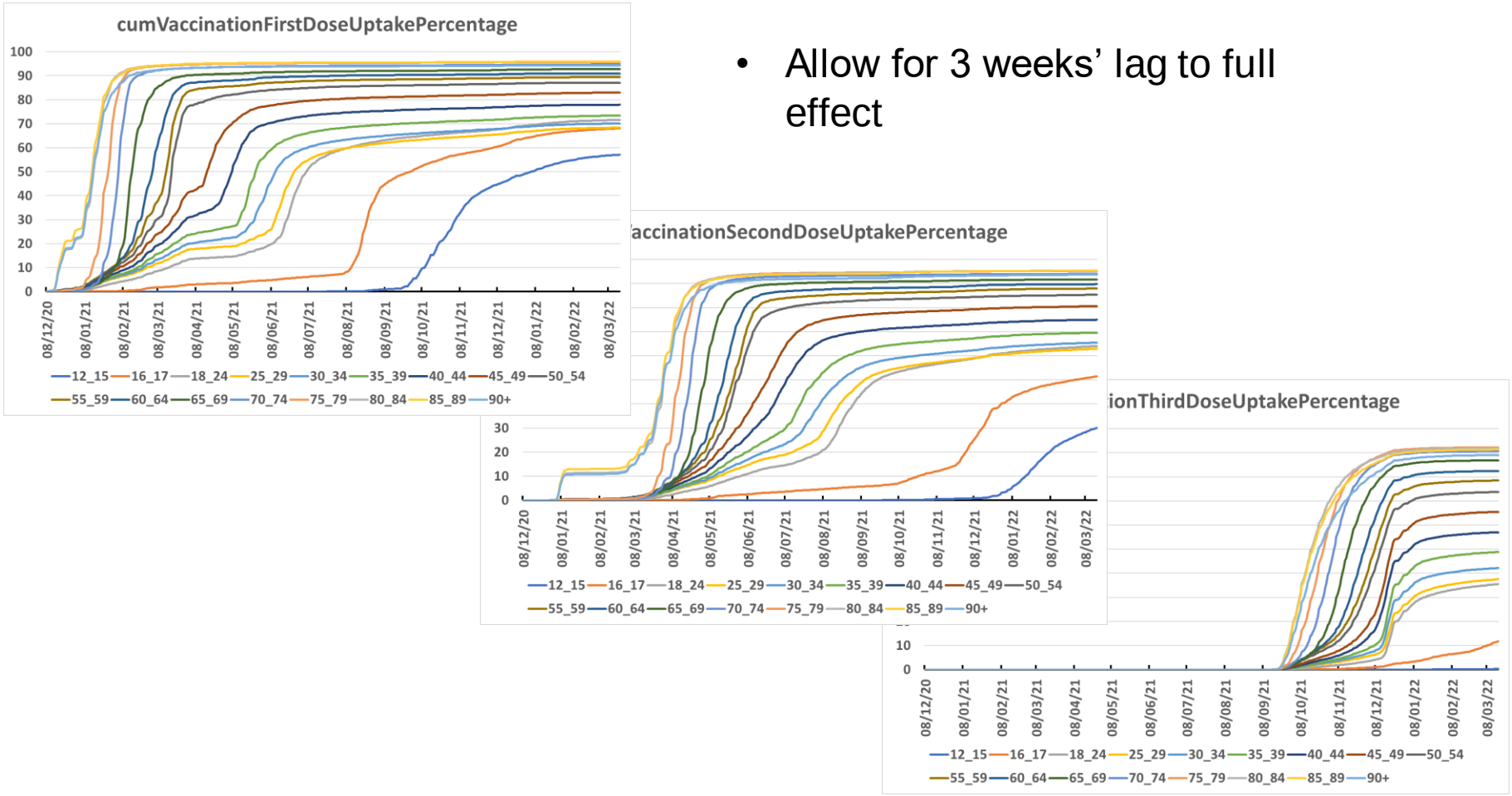
- Prevalence of variants: Sanger web site is excellent for England
  https://covid19.sanger.ac.uk/lineages/raw?lambda_type=area



- In practice, use logistic function fits for growth in variant.

2. **Case Fatality Rates from public data cont.**

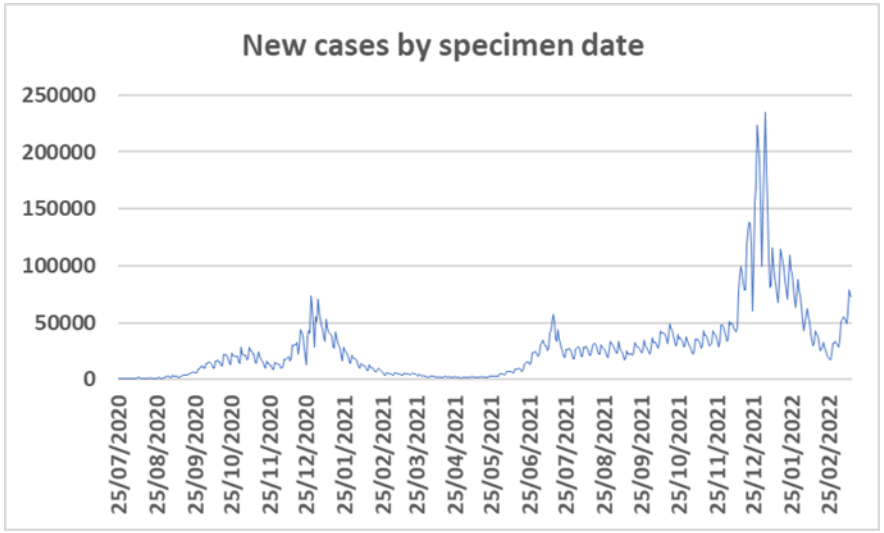
Vaccination data by age group:
https://api.coronavirus.data.gov.uk/v2/data?areaType=nation&areaCode=E92000001&metric=vaccinationsAgeDemographics&format=csv

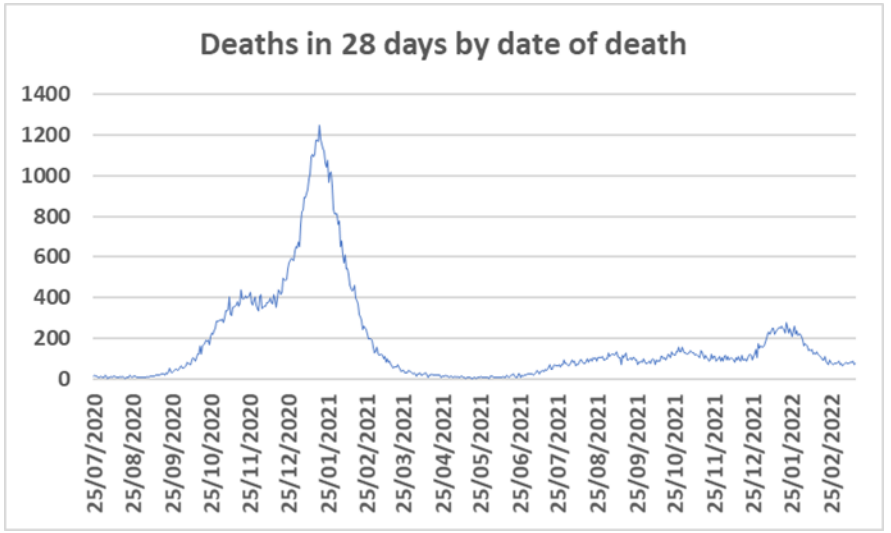


- Allow for 3 weeks' lag to full effect

2. **Case Fatality Rates from public data cont.**

- Recall:





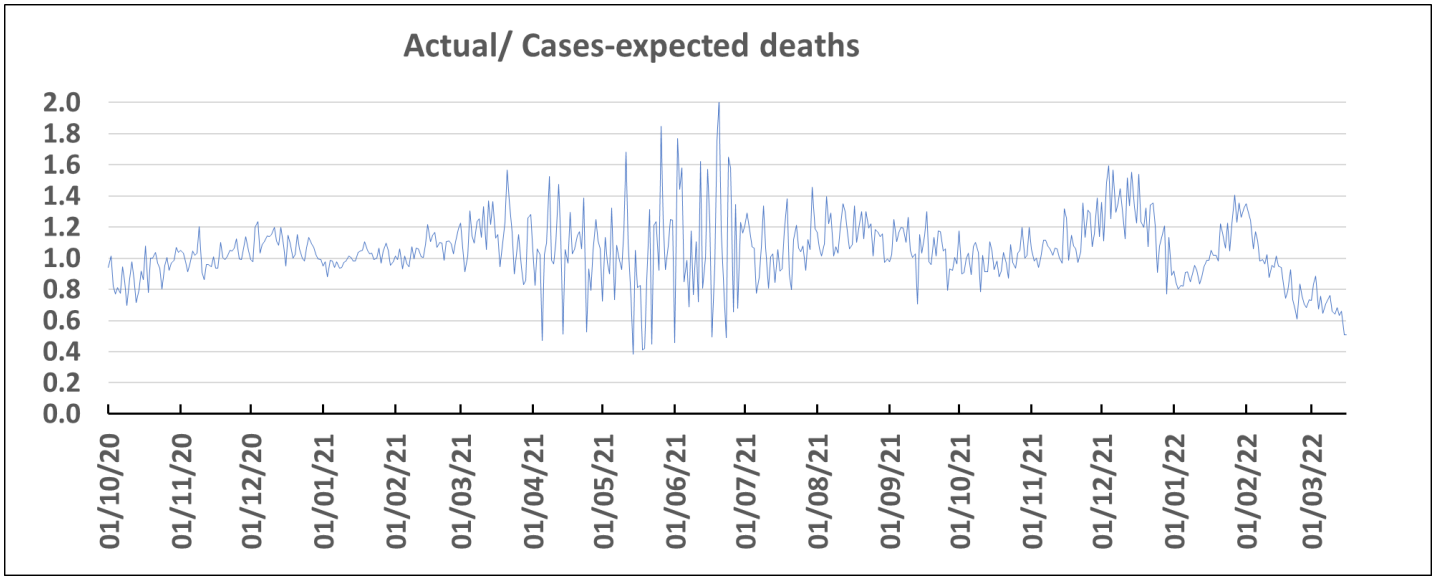
https://coronavirus.data.gov.uk/details/download

- Aim: map cases onto deaths through Case Fatality Rate (CFR)
- Copy of Excel workbook control panel shows parameters used.
- Parameters are highly correlated. Ex: in pure exponential growth, Shift ≡ Scale
- More generally, estimate for variant lethality strongly correlates with vaccine efficacy.

False positives =
B VOC =
Alpha VOC =
Delta VOC =
AY.4.2 VOC =
Omicron VOC =
Reduction first vacc. =
Reduction second vacc. =
Reduction third vacc. =

2. **Case Fatality Rates from public data cont.**

- Correlations imply lots of parameter values can produce a chart like:



Comments:

- False positives in PCR tests estimated by ONS and Oxford (summer 2020) as 0.005% - only 5 tests in 100,000 yield false positives. Sep. 2020 suggests 0.4%; above data uses FP=0. FPs may vary by prevalence the virus…
- B variant indistinguishable from base case B.1.177
- AY.4.2 (Delta variant) indistinguishable from Delta
- No attempt to disentangle Omicron variants.

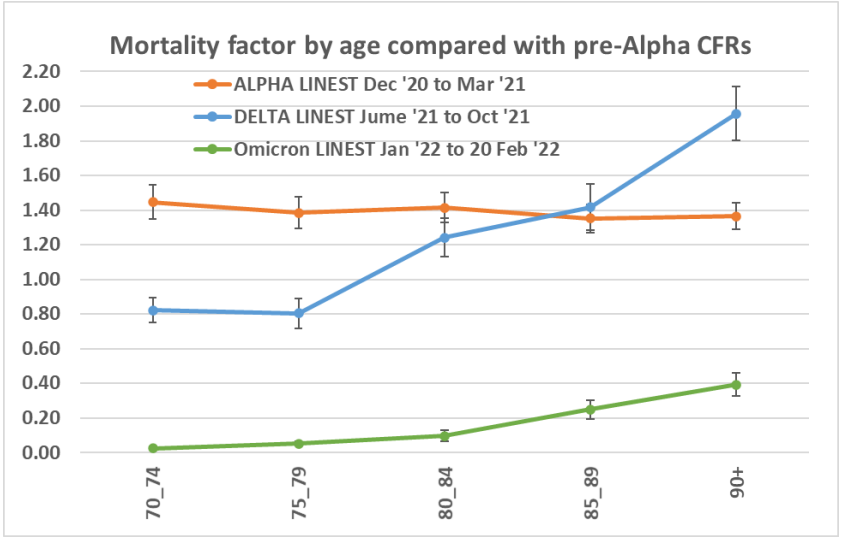
False positives =
B VOC =
Alpha VOC =
Delta VOC =
AY.4.2 VOC =
Omicron VOC =
Reduction first vacc. =
Reduction second vacc. =
Reduction third vacc. =

2. **Case Fatality Rates from public data cont.**

Dependence of other parameters on age-group

- Have seen how CFRs for Oct-Nov 2020 increase very rapidly with age. Do variant severity and vaccine efficacy also depend on age?
- To explore, **fix** vaccine effectiveness, and look at age dependence of extra variant mortality compared by age group.

- Estimates use Excel LINEST function. Age groups 70 upwards suffer >80% of deaths. Lower R^2 for younger age groups.
- Mortality even more strongly age-dependent for Delta.
- Omicron less deadly even for older age groups.
- Efficacy chosen to be 60%, 70%, 75% for 1st, 2nd, 3rd doses resp. Qualitative effect persists for any choice.



IMPORTANT: Mortality factors are after positive tests; they do not take into account reduction in likelihood of catching Covid-19 after vaccinations.

3. R-value and growth rate

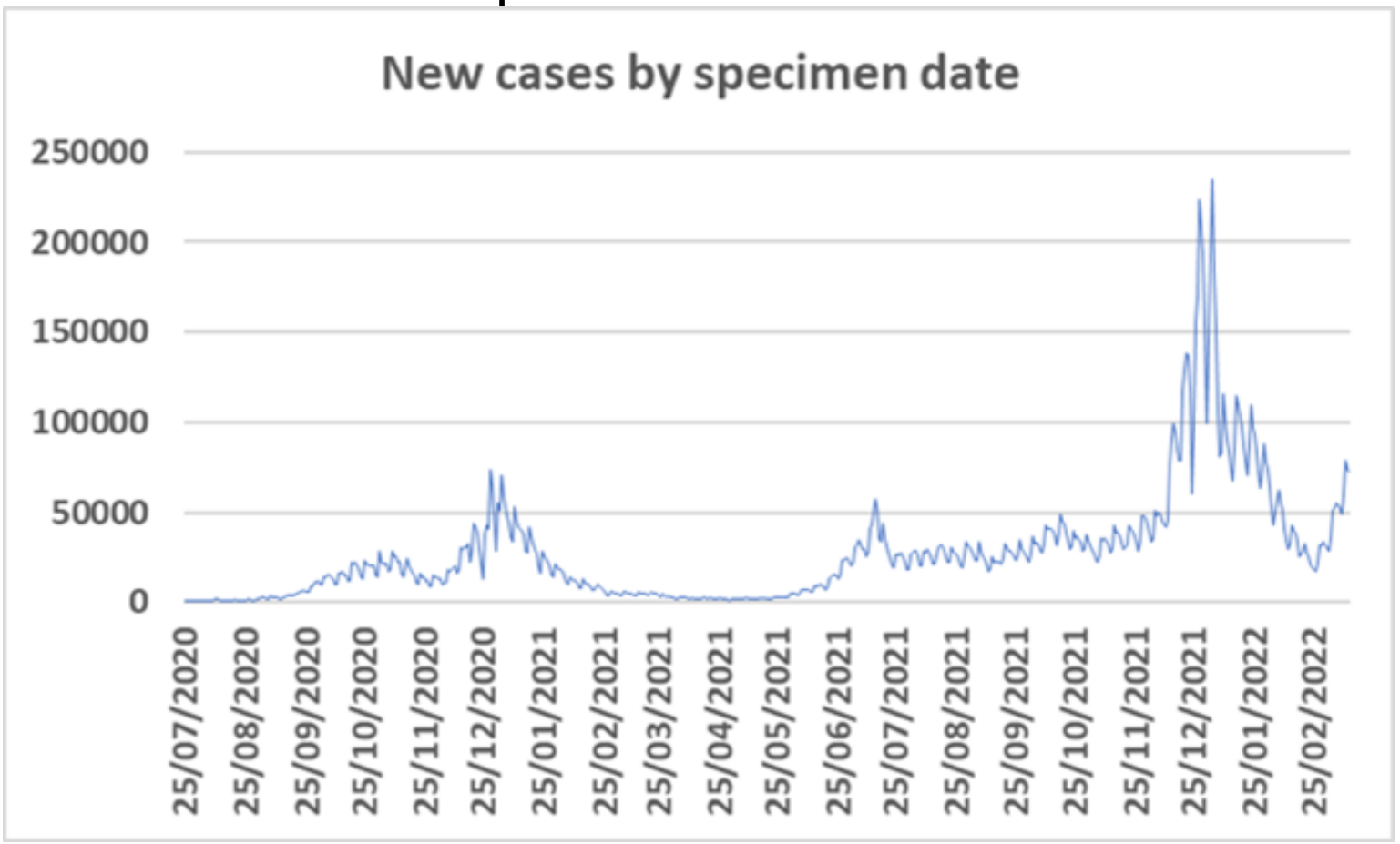
- R-value and growth rate estimates are published on Fridays https://www.gov.uk/guidance/the-r-value-and-growth-rate . The numbers are best estimates from wide range of simulations and groups, submitted through SPI-M.
- Received wide-spread publicity as indicators of the progress of Covid-19, and as important data for Government policy.
- The official figures are for *infections*. Here we look at corresponding estimates based on the publicly available data for *cases*.
- Simplest to consider first the growth rate, defined as the fractional rate of change (for us) in cases *C(t)* per day:

  $$2\,[C(t+1) - C(t)] / [C(t+1) + C(t)]$$

- Ambiguity of whether this is growth rate for day *t* or day *t+1* shifts charts by one day.
- Key point: If under-reporting of *infections* by *cases* is consistent, it is cancelled out in the ratio.

3. R-value and growth rate cont.

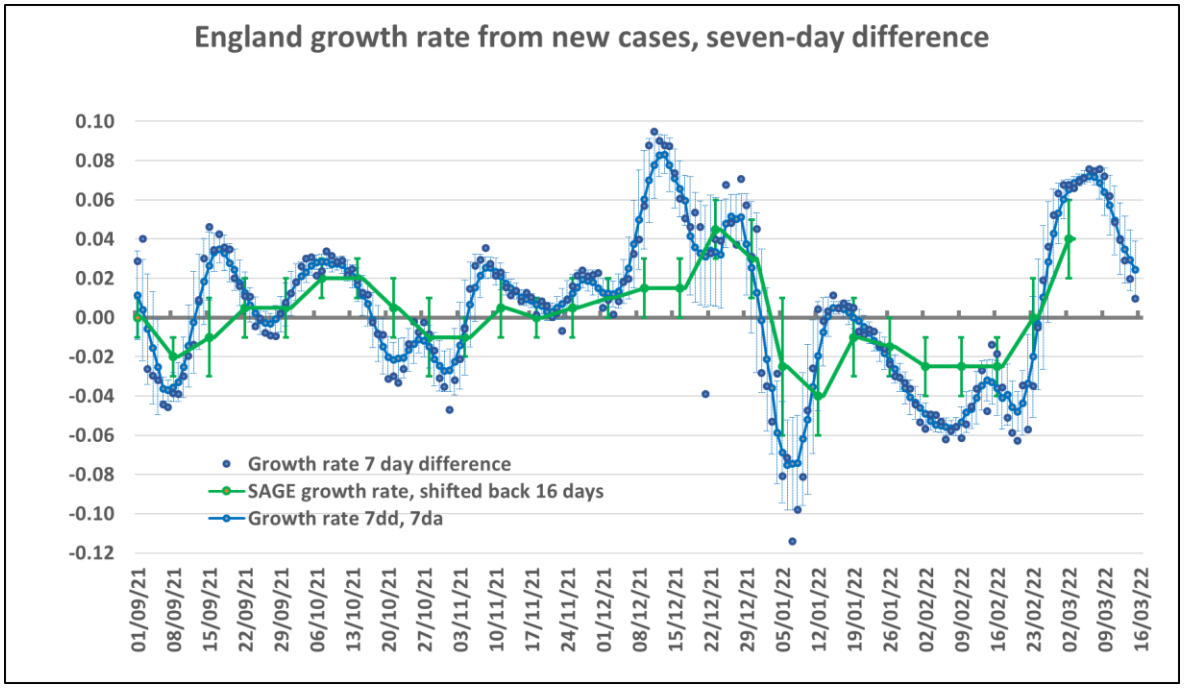
The main challenge: the statistical noise from day to day and even more obviously, the weekly variation from the lower number of tests performed at the weekend.



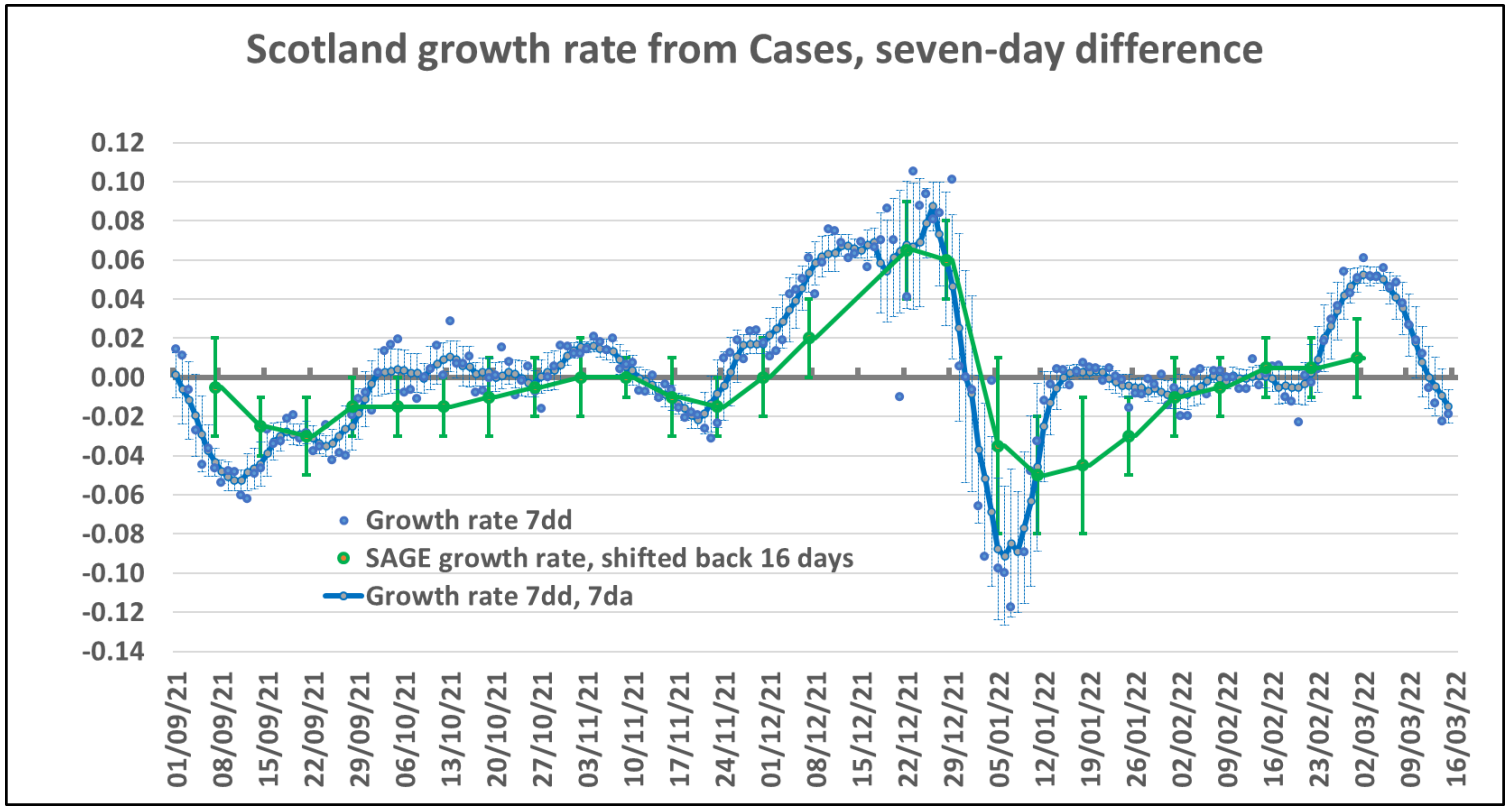
3. R-value and Growth rate cont.

- A simple approach: use 7-day difference $2\,[C(t+7) - C(t)] / [C(t+7) + C(t)]/7$
- Weekend effect is automatically eliminated, and statistical variation reduced by factor *1/7*. (GJA uses different method)
- Ambiguity of whether this is growth rate for day *t+3* or day *t+4* shifts charts by one day.

Blue dots are raw 7-day difference. Blue line is 7-day average of 7dd. Error bars are 90% CL for the 7da. Green data are weekly official estimates, *shifted back by 16 days* (GJA)
Qualitative agreement, but official estimates much smoother.



3. R-value and Growth rate cont.

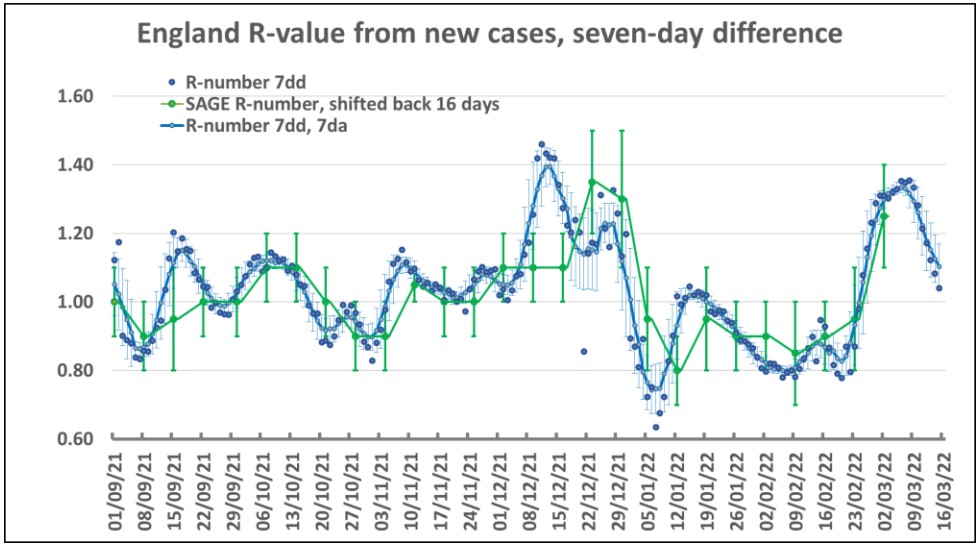


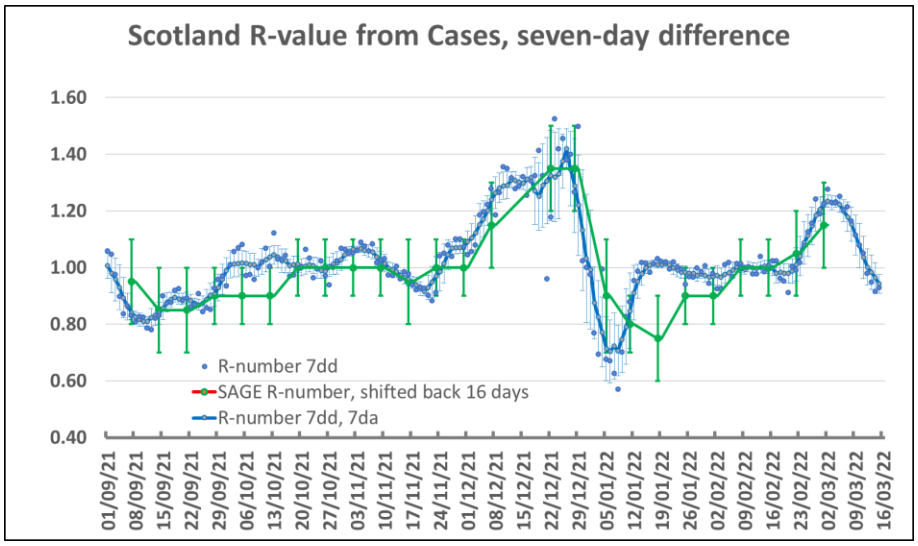
- Similar effects for Scotland. 16-day lag is now stated in the weekly announcement.
- Assuming a rapid change in numbers of new cases is reflecting a real rapid change in infections, why are official estimates of its derivative, the growth rate, so smooth?! To compare, would need to understand the averaging processes (over two weeks(?) and various research groups) used for official estimates.

3. R-value and growth rate cont.

- Similar approach for R-value (reproductive value): in simple terms, the average number of people an *infected* individual can expect to pass the coronavirus onto. We define as the average number of *cases* arising on average from one case.
- If R is constant, the change in *infections* (or, for us, *cases*) is exponential (growth if R>1, decrease if R<1), with a time constant, the generation time τ. For coronavirus, τ is estimated to be 4 to 5 days.
- As for growth rate, can overcome weekend effect by taking ratio $[C(t+7)/C(t)]^{\tau/7}$ as a simple measure of the number of cases on average resulting from 1 case. The weekend effect is cancelled and the statistical noise is reduced.
- Charts overleaf take generation time $\tau = 4$.

3. R-value and growth rate cont.

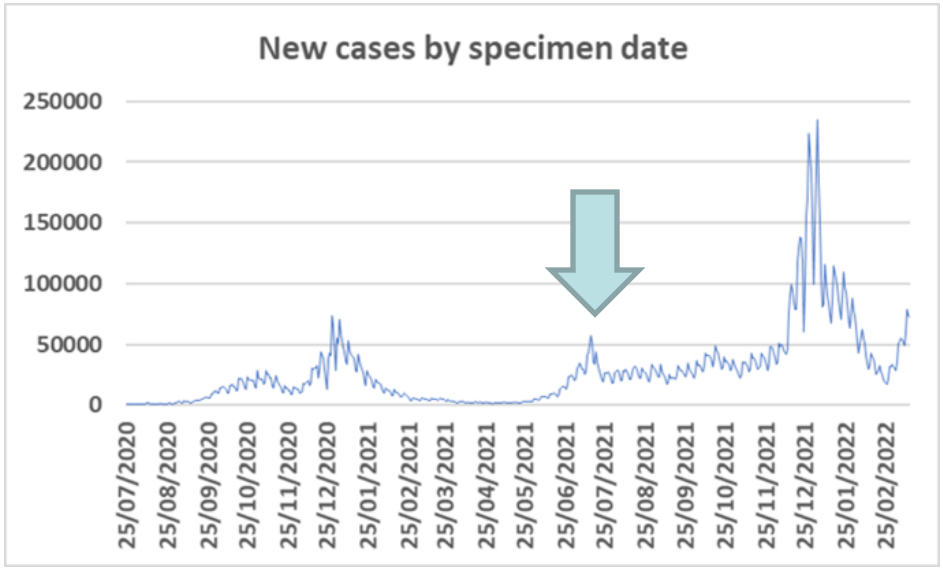


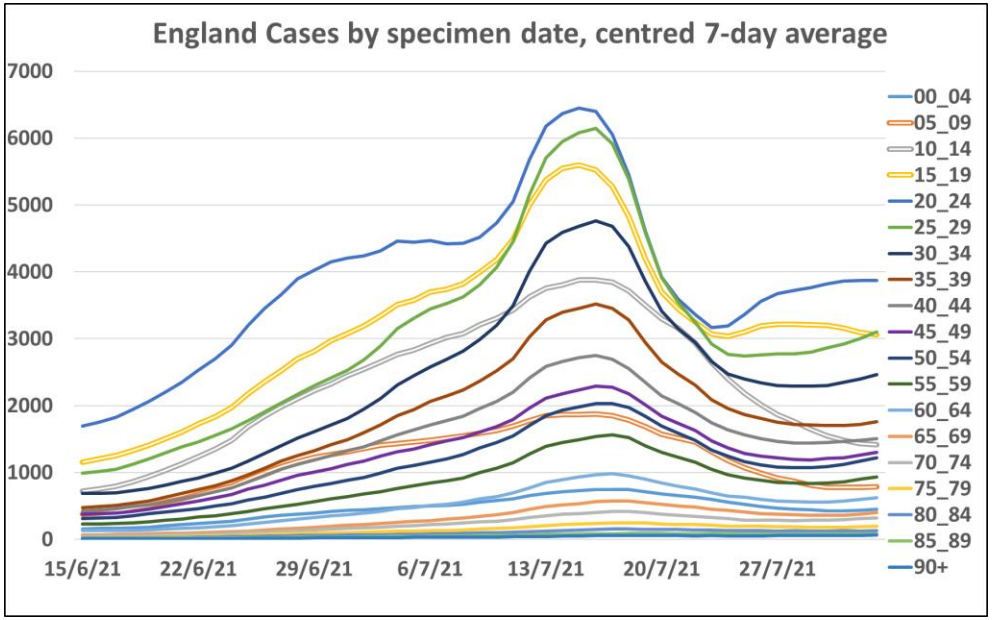


- 7-day difference is in fair agreement with official estimates, again provide the latter are interpreted as 16 days out of date.
- Whitty (ack. to GJA): What makes an academic paper useful for health policy? | BMC Medicine | Full Text (biomedcentral.com). "Since the policy process tends to be very fast, papers must be timely. An 80% right paper before a policy decision is made is worth ten 95% right papers afterwards, provided the methodological limitations imposed by doing it fast are made clear."
- How useful to policy makers are R-value and growth rate estimates which are 16 days out of date when published?!

4. Euro2020

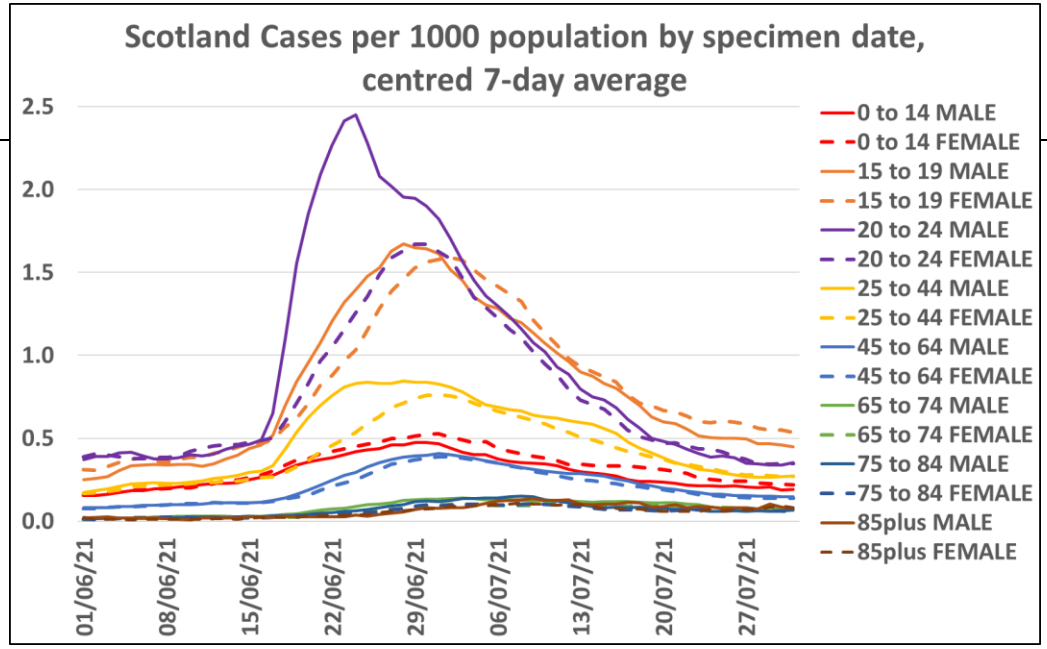
- Back to cases data.
- Look in detail at peak in England data in June/July 2021.
- Age demographics show peak dominated by increase and decrease in age groups 15-19 to 55-59, broader and increasingly delayed for older age groups.
- England first game 13/6, final 11/7. Peak a few days after.

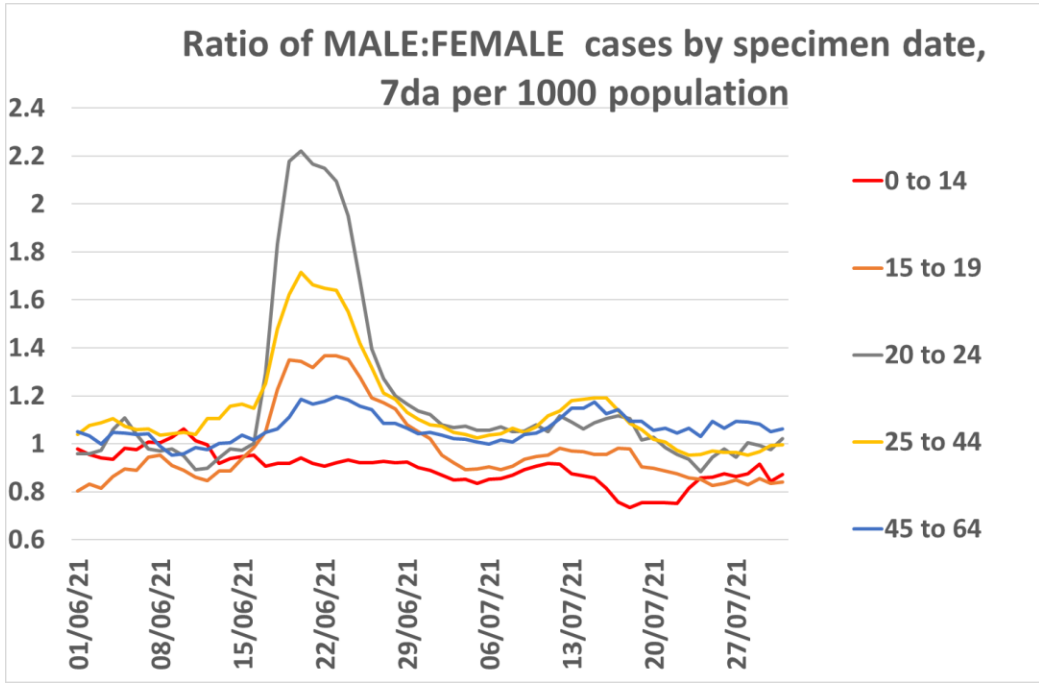




4. Euro2020 cont.

- Scotland data. Age groups unhelpful, hence per 1000 pop.
- Scotland first game 14/6, then v. England 18/6, last game 22/6.
- Significant relaxation of restrictions in pubs etc during Euro2020.
- Peak most prominent in 20-24, 15-19 and 25-44 males (solid lines), who spread infection quite rapidly into female contemporaries. Peak occurs 4 days after England game.
- Effect seen most sharply in ratio M:F (thanks to GJA)
- Secondary peak after final on 11/7?!
- Illustrates dynamic of rapid spread within highly connected subgroups and slower spread across groups.





Conclusions

- **MS Excel is a useful tool for this kind of modelling.**
  - **Very easy to get into and can build quite complex models**
  - **I found the transparency of data helpful**
  - **Charts are a useful check**
  - **Made extensive use of Solver, LINEST and other tools. Excel has more than 300 built in functions.**
- **BUT**
  - **I did a lot of stuff by hand – data input and update, and cut and paste.**
  - **I could probably have done it more easily with experience, and if I had known some Visual Basic for Applications (VBA)**

Acknowledgements

- **David Spiegelhalter (Cambridge)**
- **Graeme Ackland (Edinburgh)**
- **James Ackland (Cambridge)**
- **Rowland Kao (Edinburgh)**
- **Chris Bishop, Andy Gordon and Tom Minka (MS Research Cambridge)**
- **Conversations with many others.**

THANK YOU!