

Official estimates of community Covid positivity: estimating rare occurrences using multi-level regression with post-stratification in R

Dr Stephanie Allen

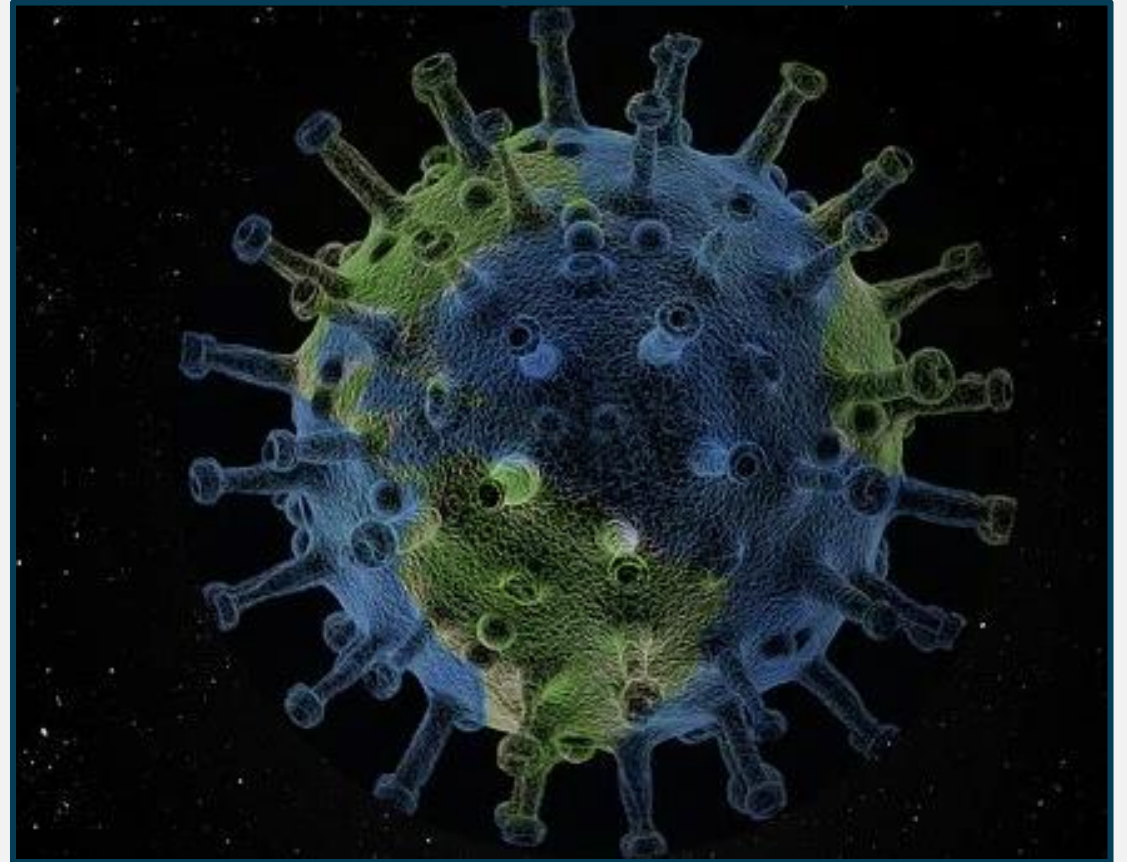
Senior Analyst

COVID-19 Infection Survey analysis division

Health Analysis and Pandemic Insight

COVID-19

- The coronavirus (COVID-19) pandemic has impacted communities and economies around the globe
- Taking centre stage in government priorities, there was an urgent requirement to track positivity rates across the UK so that informed decisions could be made in a timely manner
- To achieve this the COVID-19 Infection Survey was set up as a monitoring program across England, Wales, Northern Ireland and Scotland



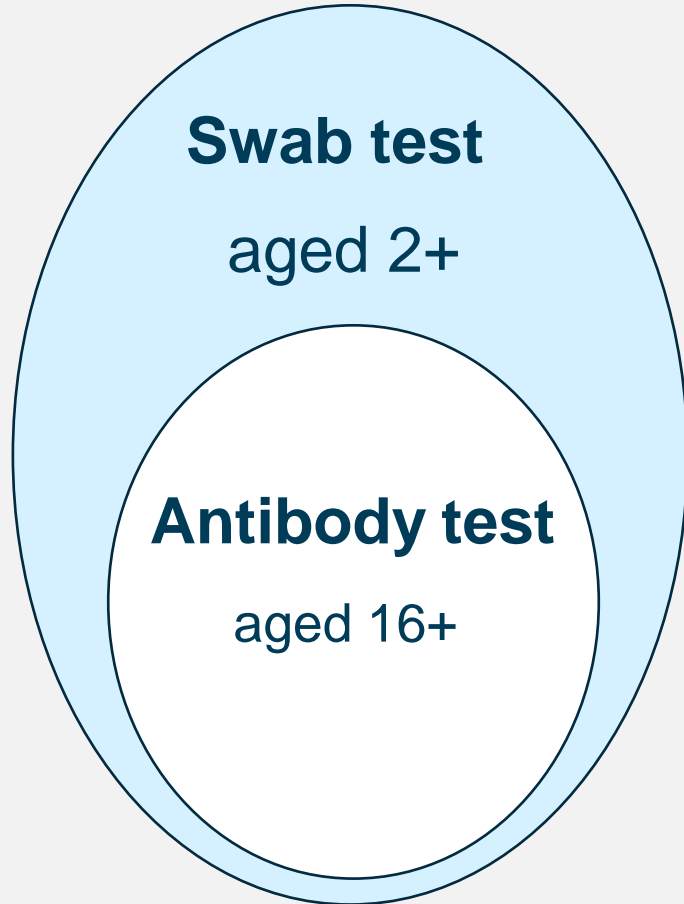
Covid Infection Survey

Representative household survey

- **Positivity and Incidence (Swab survey):** aged 2+ living in private households - tested with PCR tests each week for the 5 weeks and then every month thereafter - estimates COVID-19 with or without symptoms
- **Antibody survey:** aged 16+ living in private households; subset of participants are selected for blood tests for COVID antibodies
- Following a pilot, in July 2020 we expanded the survey, to test nearly **200,000 people per two weeks across the UK**



Every month we collect:



Demographics

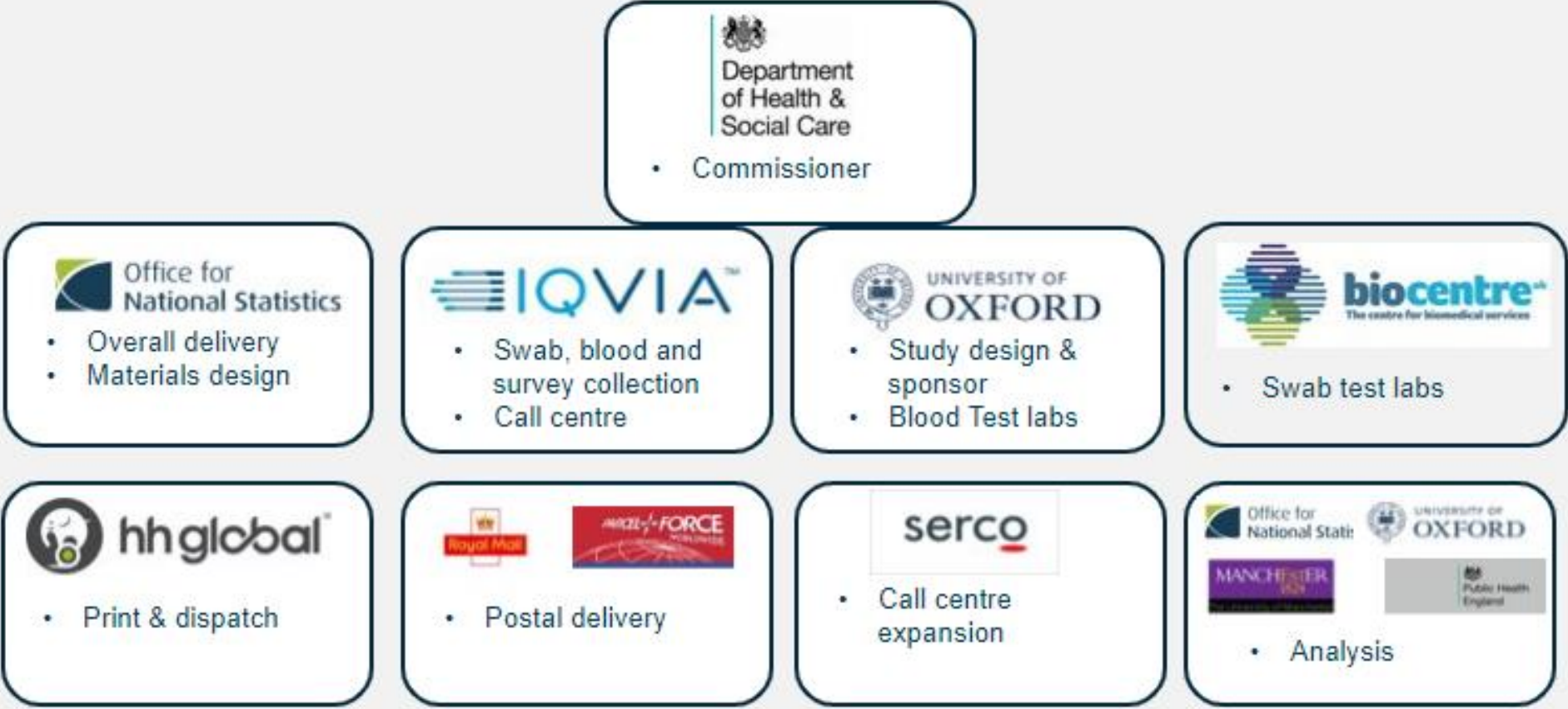
Occupation and Sector

Contact with others

Behaviour



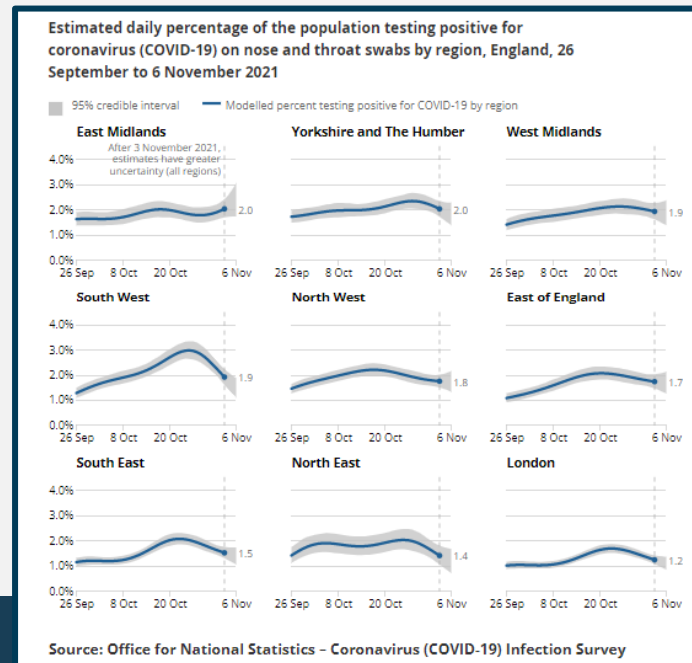
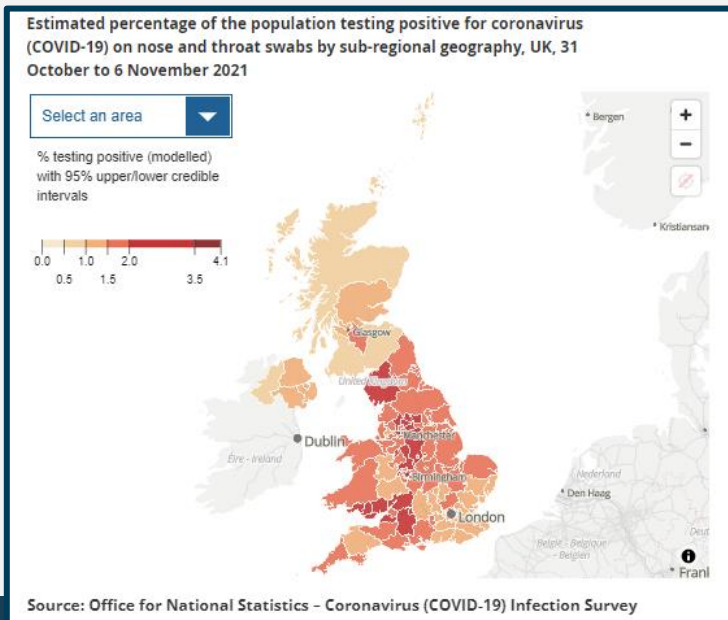
We work in partnership



NB: IQVIA now have multiple field partners from across the survey and medical research industry, inc Ipsos MORI, Natcen, Kantar

Statistics produced in


- Daily positivity rates
- Positivity estimates by region and age group in England
- Sub-regional positivity estimates for the UK.



Estimated percentage of the population testing positive for coronavirus (COVID-19) on nose and throat swabs, UK, 8 November 2020 to 6 November 2021

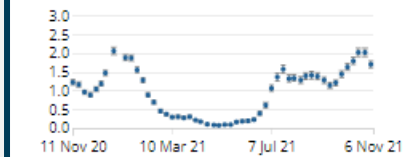
Official reported estimates of the rate of COVID-19 infections in private households in England, Wales, Northern Ireland and Scotland.

Modelled estimates are used to calculate the official reported estimate. The model smooths the series to understand the trend and is revised each week to incorporate new test results.

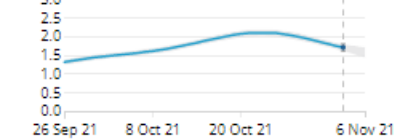
 Estimates with 95% confidence/credible intervals

England

Percentage testing positive for COVID-19

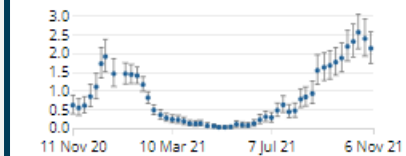


Percentage testing positive for COVID-19
After 3 November 2021, estimates have greater uncertainty

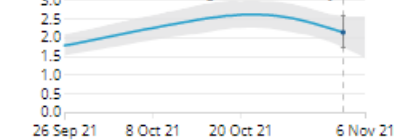


Wales

Percentage testing positive for COVID-19

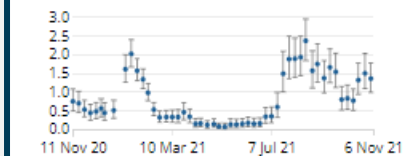


Percentage testing positive for COVID-19
After 3 November 2021, estimates have greater uncertainty

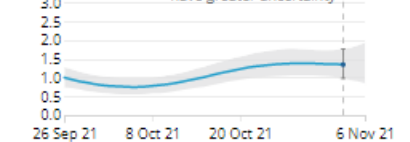


Northern Ireland

Percentage testing positive for COVID-19

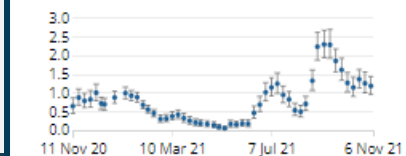


Percentage testing positive for COVID-19
After 3 November 2021, estimates have greater uncertainty

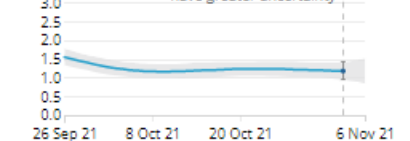


Scotland

Percentage testing positive for COVID-19



Percentage testing positive for COVID-19
After 3 November 2021, estimates have greater uncertainty



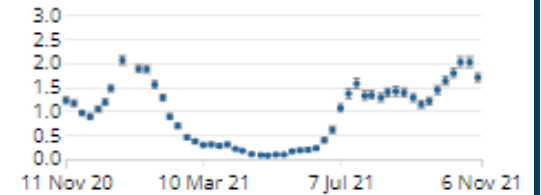
Source: Office for National Statistics - Coronavirus (COVID-19) Infection Survey

Why we use

- R is a free and Open Source, making official statistics packages well shared, developed and tested
- R-based software is exchanged through strictly enforced technical standards – ensuring quality when we use it
- High software and community support, R provides lots of opportunities to find solutions online so we can continue to develop our methods to the highest quality
- It also makes our methods re-producible, providing the opportunity to build in quality checks which are vital when producing official estimates
- Excellent community of users

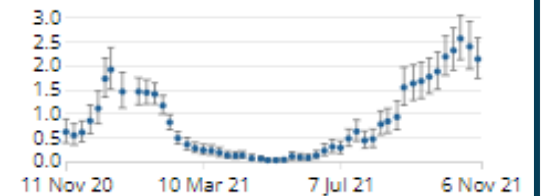
England

Percentage testing positive for COVID-19



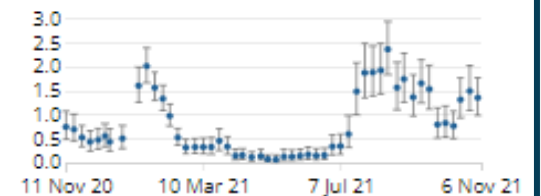
Wales

Percentage testing positive for COVID-19



Northern Ireland

Percentage testing positive for COVID-19



Scotland

Percentage testing positive for COVID-19



Source: Office for National Statistics – C

Producing official estimates

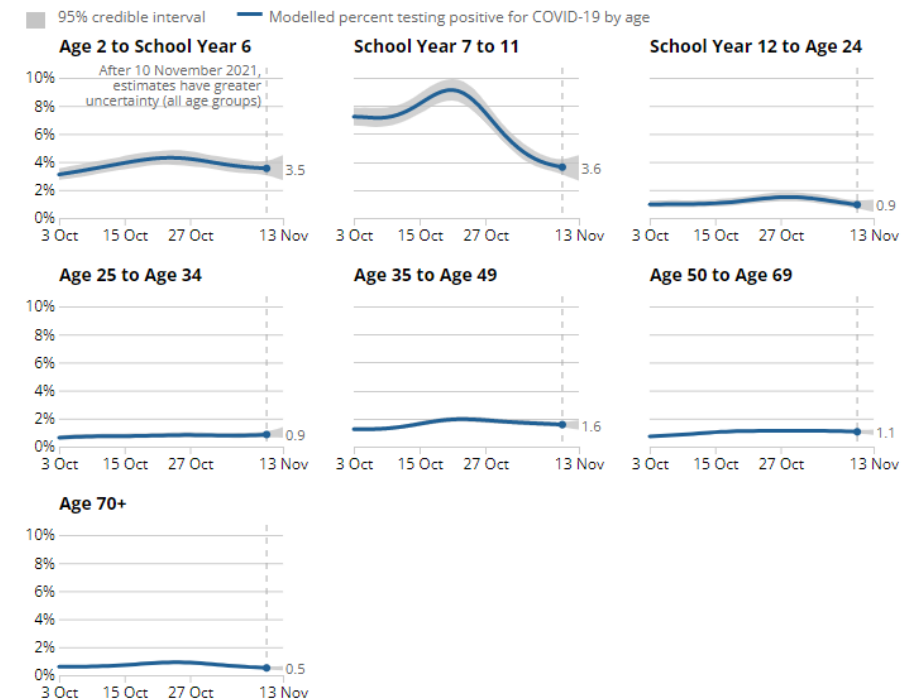
- Early in the pilot, it became clear that a fortnightly weighted estimate was too blunt to understand what was happening in the pandemic.
- Understanding what was happening day-by-day, particularly whether the trend was increasing or decreasing, was vital.
- Although we collect a lot of data everyday, fieldwork targets are met over a month, so needed a method that could produce a robust, representative, daily estimate.

Solution > Enter R code and the use of multi-level regression with post-stratification

- Developed by our academic research partners at the University of Oxford and the University of Manchester, and produced by the ONS.

Figure 4: The percentage of people testing positive decreased in most age groups in the week ending 13 November 2021

Estimated daily percentage of the population testing positive for coronavirus (COVID-19) on nose and throat swabs by age group, England, 3 October to 13 November 2021

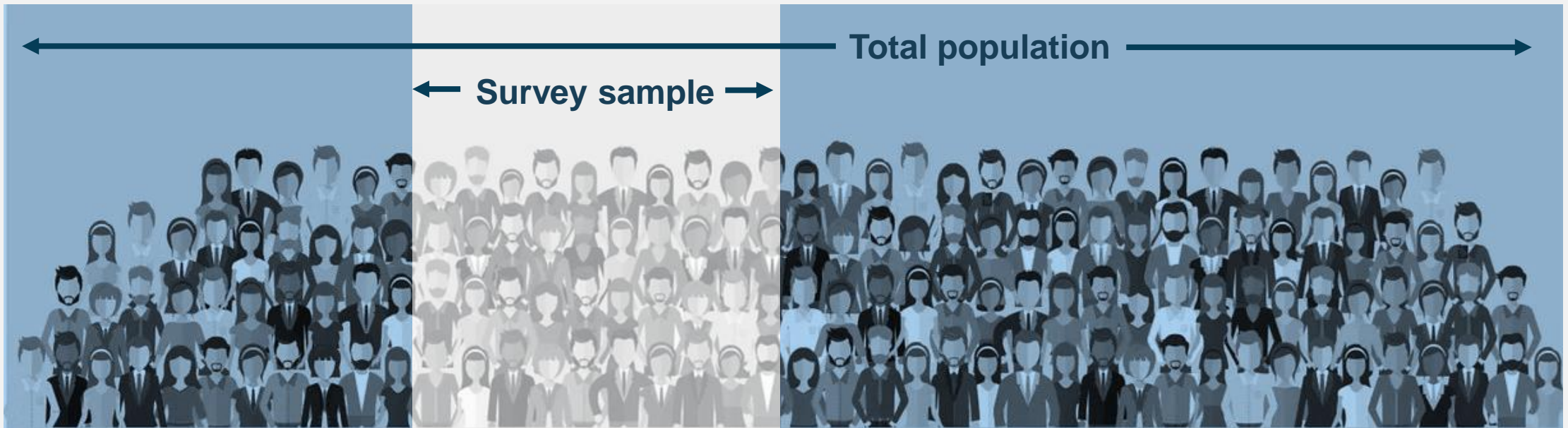


Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

Details about the methodology are also provided in the peer-reviewed paper from our academic collaborators published in the [Lancet Public Health](#).

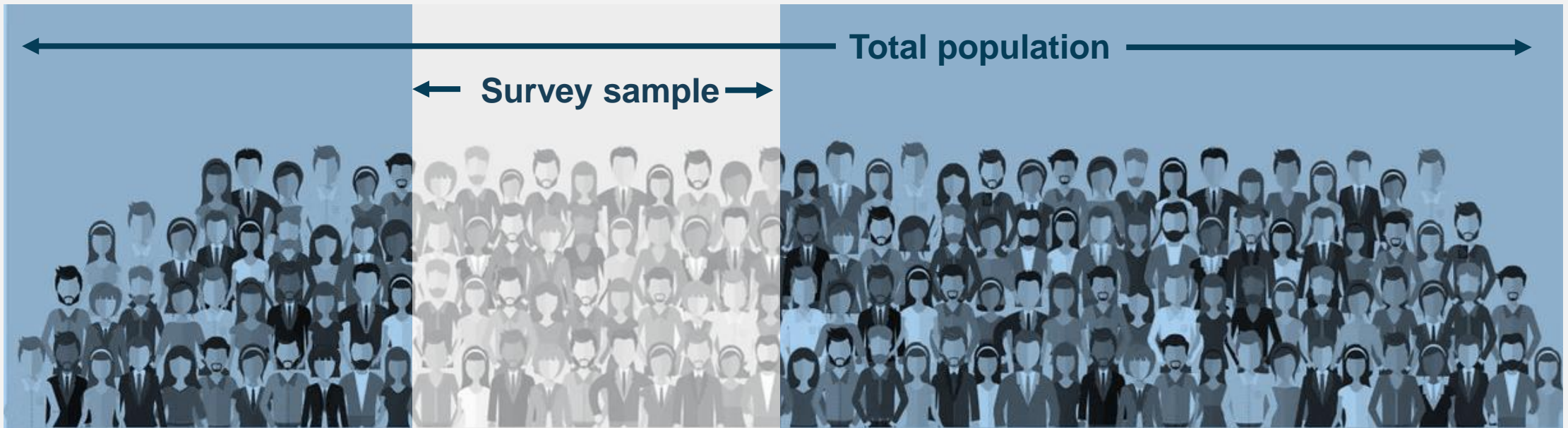
What is MRP?

- Multilevel regression and poststratification (MRP) is a statistical technique used for correcting model estimates for known differences between a sample population (data collected in the covid infection survey), and a target population (a population you would like to estimate for e.g. the UK)



What is MRP?

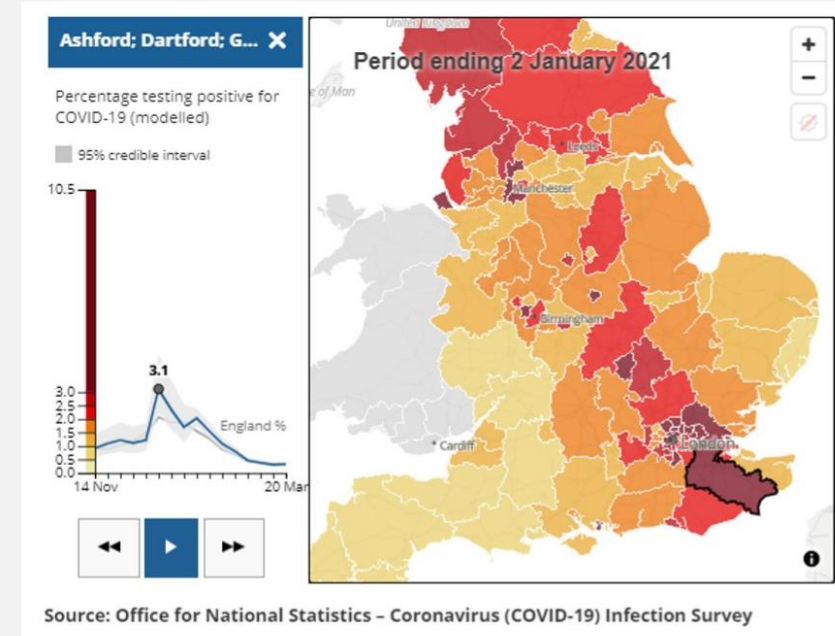
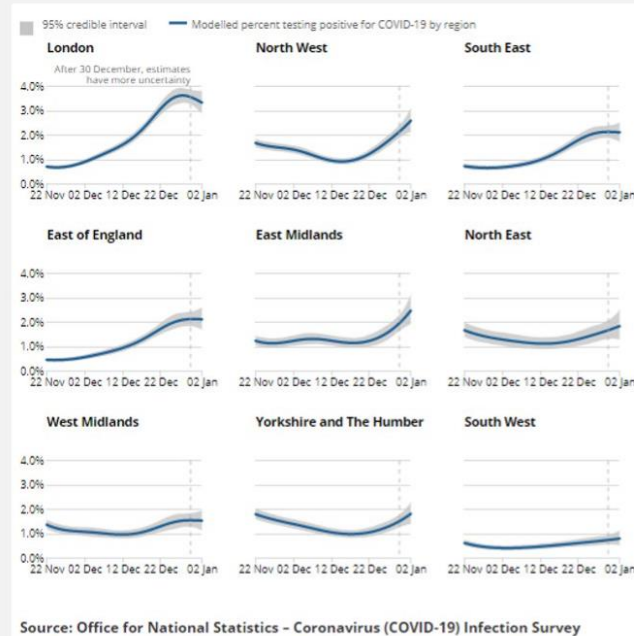
- Multilevel regression - generates a robust estimate by using what has happened before and after, as well as in nearby categories (e.g. spatial proximity, categorical proximity). Very helpful when measuring something rare. It can also smooth noisy estimates or where there is too little data
- Poststratification - the process of adjusting the estimates, to meet the attributes of the total population essentially weighted average of estimates



The end result

- This technique means that the covid infection survey data estimates the covid positivity rates across the whole UK
- While being able to define patterns by region and age groups, which helps to inform policy decisions

Analysis to inform local lockdowns



MRP in R

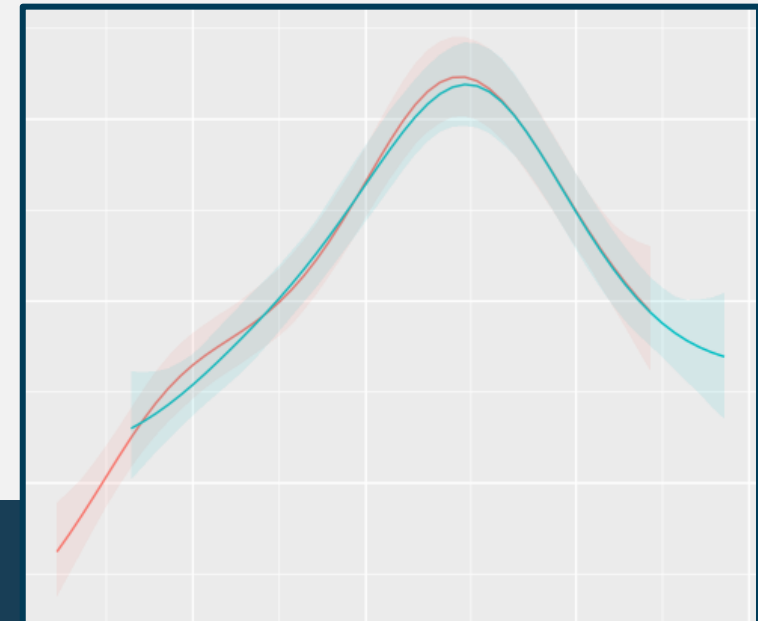
- R is an excellent tool for running big models like MRP
- We can use simple code to monitor our run times
- We can use if and else if functions to determine where our output are saved depending on the data type = saving time post-model
- We can produce QA aspects to the code – for example outputting statistical information such as Akaike information criterion (AIC) to compare iterations of the model for best fit and overall model runs between dates so we can assess how the model fits, as more information from the survey comes in

```
start_time <- sys.time()

----- mrp model -----

end_time <- sys.time()
mod2runtime <- end_time - start_time
```

```
if (variant_type == 1){
  tables_directory <- "outputs/tables/variant1"}
else if (variant_type == 2){
  tables_directory <- "outputs/tables/variant2"}
else if (variant_type == 3){
  tables_directory <- "outputs/tables/variant3"}
else if (variant_type == 4){
  tables_directory <- "outputs/tables/variant4"}
else{print("INCORRECT VARIANT TYPE ENTERED")}
```



The data we've produced

Estimated percentage of the population testing positive for coronavirus (COVID-19) on nose and throat swabs, UK, 7 to 13 November 2021

Country	Estimated average % of the population that had COVID-19	95% credible interval		Estimated average number of people testing positive for COVID-19	95% credible interval		Estimated average ratio of the population that had COVID-19	95% credible interval	
		Lower	Upper		Lower	Upper		Lower	Upper
England	1.51	1.42	1.60	824,900	775,500	873,700	1 in 65	1 in 70	1 in 60
Wales	1.84	1.47	2.24	55,800	44,800	68,100	1 in 55	1 in 70	1 in 45
Northern Ireland	1.53	1.16	1.96	28,100	21,300	35,900	1 in 65	1 in 85	1 in 50
Scotland	1.06	0.86	1.29	56,000	45,500	68,100	1 in 95	1 in 120	1 in 75

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

Estimated percentage of the population testing positive for coronavirus (COVID-19) on nose and throat swabs, UK, up to 13 November 2021

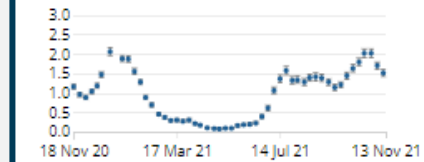
Official reported estimates of the rate of COVID-19 infections in private households in England, Wales, Northern Ireland and Scotland.

Modelled estimates are used to calculate the official reported estimate. The model smooths the series to understand the trend and is revised each week to incorporate new test results.

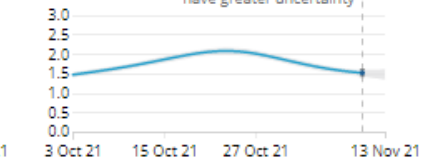
Estimates with 95% confidence/credible intervals

England

Percentage testing positive for COVID-19

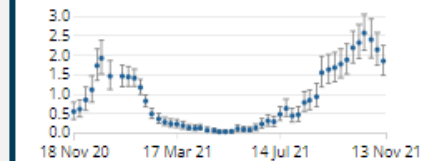


Percentage testing positive for COVID-19
After 10 November 2021, estimates have greater uncertainty

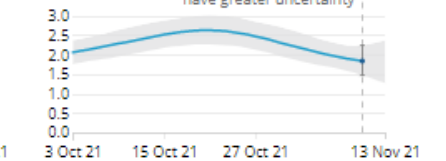


Wales

Percentage testing positive for COVID-19

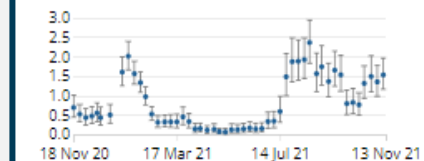


Percentage testing positive for COVID-19
After 10 November 2021, estimates have greater uncertainty

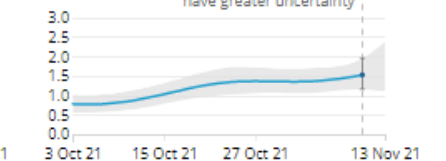


Northern Ireland

Percentage testing positive for COVID-19

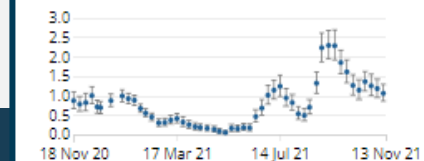


Percentage testing positive for COVID-19
After 10 November 2021, estimates have greater uncertainty

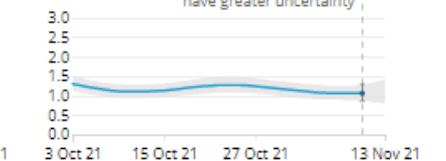


Scotland

Percentage testing positive for COVID-19



Percentage testing positive for COVID-19
After 10 November 2021, estimates have greater uncertainty



Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey

What we've learned about

- Upfront time improving coding is time well spent, saves time later
- If you're running repeated analysis - make it as simple as possible to change critical variables without risking unintentional code changes
- Add code to monitor run-times if you're running lots of CPU intensive, time-critical analysis
- Markdown is a brilliant tool for reporting on quality, for analytical producers as well as for sharing with internal users to help aid analysis.

Summary

- The MRP allows for the ONS to create official estimates accurately across the UK demonstrating changes in different regions and across different age groups
- The MRP allows the ONS to deal with relatively small numbers of positive tests and ensure estimates are unbiased
- The use of statistics and R software supports decisions regarding the continued need for control measures to contain the spread of COVID-19 rely on accurate and up-to-date information about the number of people and risk factors for testing positive



Thank you

Dr Stephanie Allen

stephanie.allen@ons.gov.uk

Senior Analyst

COVID-19 Infection Survey analysis division

Health Analysis and Pandemic Insight