Covid-19 Infection Survey at ONS

IASS Webinar - 3 September 2020

Salah Merad

Methodology Division

Office for National Statistics (ONS)

Salah.merad@ons.gov.uk

Background

- A collaboration between ONS, DHSC, University of Oxford and University of Manchester

- Phase 1 of Covid Infection Study: a large pilot

- Aim is to measure prevalence, incidence of Covid-19 (swab test), and percentage of people already infected (antibody test) in the community in England

- Data collection started on 26 April

- Ethics approval
  - Initially only households who consented to follow-up could be contacted
  - Has been extended to all households

Sample design

- Population:
  - Prevalence and Incidence (Swab survey): People aged 2+ living in private households
  - Antibody survey: People aged 16+ living in private households

- Initial issued sample
  - Follow-up to ONS surveys (LFS, Opinions and Lifestyle): unclustered design
  - Total of 20275 households were invited to take part in the "Swab" survey
  - A subsample of about 4000 households invited for Antibody test
  - Limited to England

- Enhancement follow-up samples of about 5000 households each being issued weekly since 27 May: Phase 2 of the study

- Extended to Wales and Northern Ireland

- Additional samples drawn directly from Address Register – oversampling in hotspots but unclustered

- In total, about 400000 households have been invited so far

- Households approached for 5 consecutive weeks and then monthly (16 visits in total)

Data collection

- Households are asked to ring a call centre
- Nurses visit households
  - Swab kits delivered by a nurse (Swab only subsample): self-administered
  - Nurses with PPE take blood samples and supervise swab taking
- Swab and blood samples analysed by designated laboratories
- Households fill out an online questionnaire: age, sex, ethnicity, work status, job title, responsibilities, whether working in health care, social care, away from home, symptoms, contacts with infected people, shielding, isolating, travel abroad
- IQVIA was contracted out to run the data collection
- Incentives : £50 per person at first visit, and £25 per person from 2nd visit

Data processing

- Data were sent to ONS daily initially, but now it's twice a week
- Data are captured in several files that need linking
- Processing of data is now automated using stata syntax
- Several datasets are produced: non-overlapping 14-day swab datasets; all responses swab dataset (ever/never); incidence base and follow-up datasets; antibody datasets (ever/never, 14-day)

- In June, we achieved about 9200 households out of about 20000 in issued sample
- Response rate is much lower in samples that are not follow-up to ONS surveys

Weighting

- Prevalence
  - Calibration
    - Age by sex; Region (population projections)
    - Household size (Distribution in Annual Population Survey 2019)
      - To correct for potential attrition and non-response bias unaccounted by age, sex and region

- Incidence
  - Calibration of base dataset
  - Attrition model between base and incidence dataset: age by sex, region, ethnicity, household size, type of occupation (health/social care public-facing, other)

Estimation of prevalence

- Direct weighted estimates at national level but not for breakdowns: they lead to unstable estimates because of low number of positive results

- Model-based estimates (MRP method and trend estimation)
  - Currently used to report headline estimates of prevalence at national and regional levels

Trend Estimation using Dynamic multi-level regression and post-stratification (MRP) (by Koen Pouwels, Oxford Uni)

- MRP introduced by Gelman and Little (1997) and increasingly used for adjusting survey samples to be more representative (e.g. accurate predictions of Brexit by YouGov)

- Step 1: multilevel regression model generating outcome (e.g. prevalence) of interest as a function of socio-demographic and geographic variables.
    - We used a Bayesian multilevel generalised additive model with complementary log-log link and time modelled using thin-plate splines, allowing for an interaction between time and region.

- Step 2: Resulting outcome estimates for each respondent type are post-stratified by the percentage of each type in the overall population

Estimated percentage of the population in England testing positive for the coronavirus (COVID-19) on nose and throat swabs since 26 April 2020
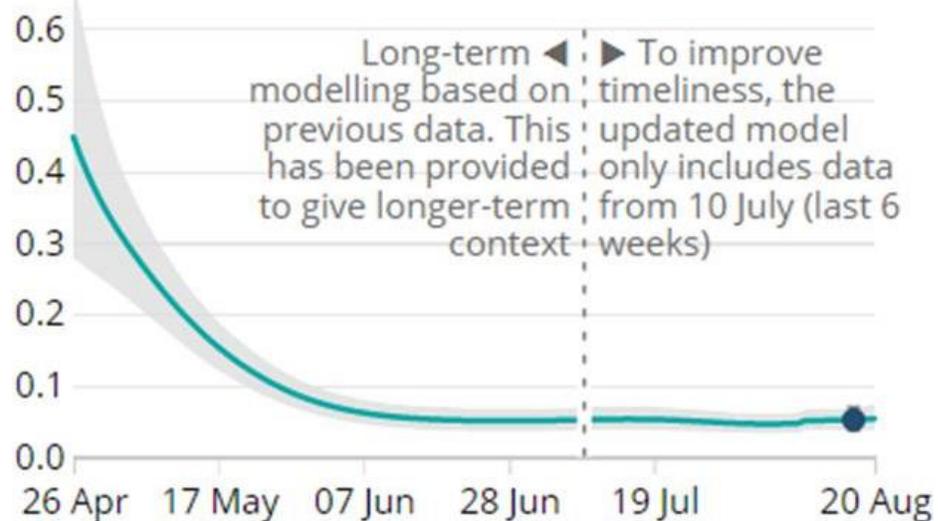
Official reported estimates of the rate of COVID-19 infections in the community in England.
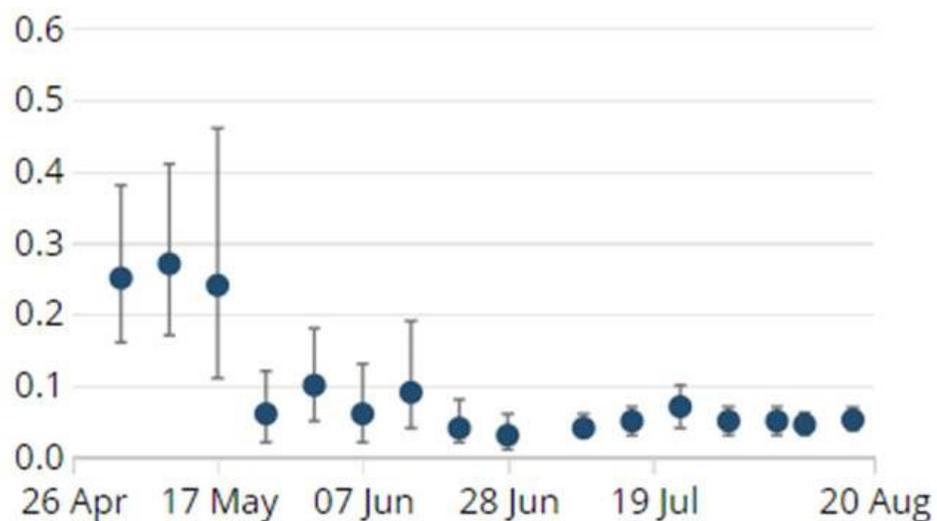
Modelled estimates are used to calculate the official reported estimate. The model works by smoothing the series to understand the trend and is revised each week to incorporate new test results.

Percentage testing positive for Covid-19

Percentage testing positive for Covid-19

Long-term ◄ : ► To improve modelling based on : timeliness, the previous data. This : updated model has been provided : only includes data to give longer-term : from 10 July (last 6 context : weeks)
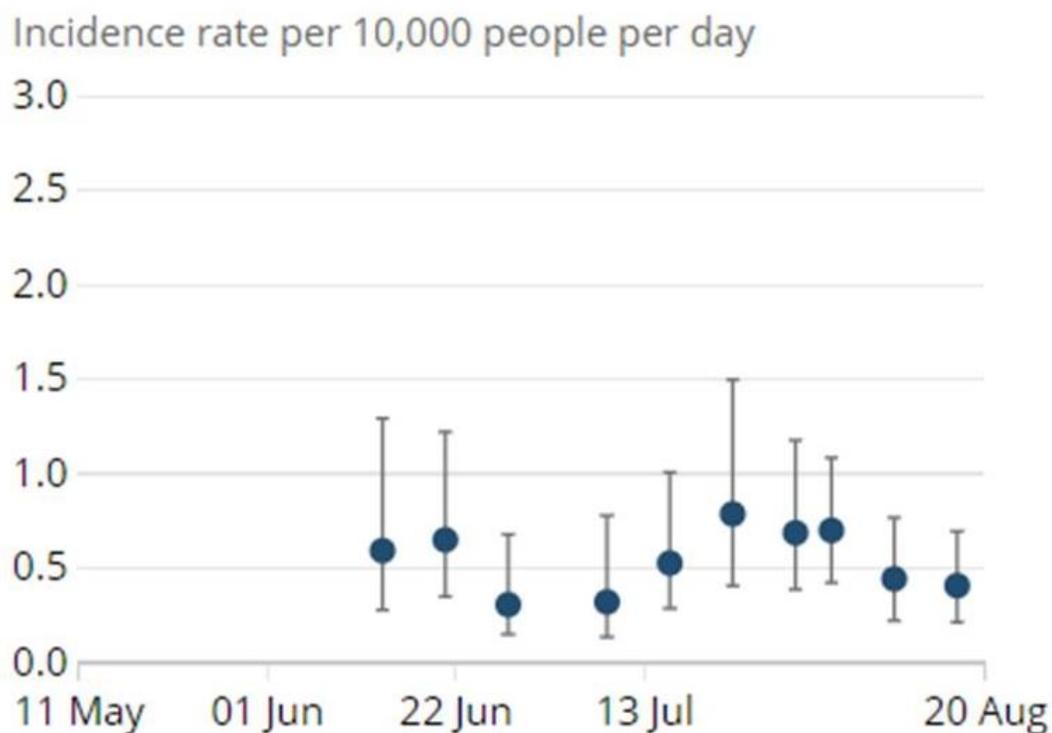
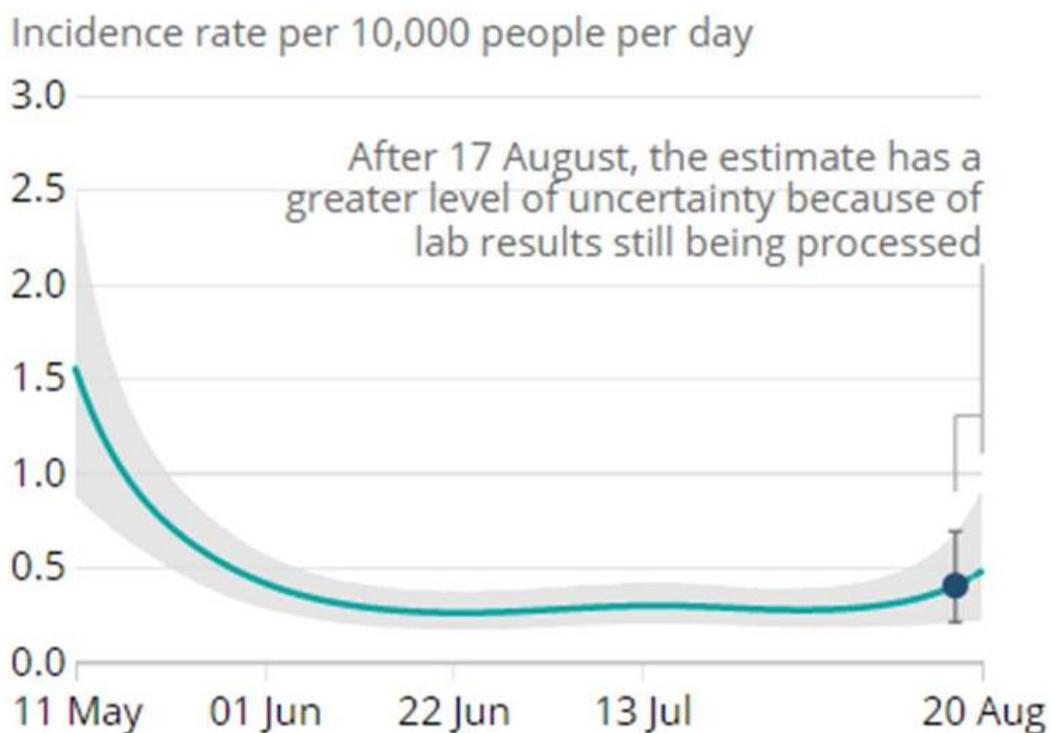**Source: Office for National Statistics**

Incidence (by Sarah Walker, Oxford University)

- Ratio of number of cases of a disease (here new COVID-19 infections) to the total time the population being monitored is at risk
  - Exclude individuals already positive on their first swab test – not "at risk" for a first infection – although will look at clearance/re-infection separately
  - Data are "interval-censored" – we do not know the time a person became infected, only the dates of their last negative and first positive swab
  - Use midpoint imputation (take time of event to be midpoint between these two dates) – but some people on monthly visits and most positives are positive only once – therefore restrict the event date to be no more than 7 days before the last positive
  - Denominator is the sum of the time from the first negative swab test in the study to the minimum of the last negative swab test (non-case) and this event time (case)

- Analysed using Poisson models using the MRP framework, because there is no "natural" time zero so time-to-event/Cox models – time zero is just when they joined the study

- Issues moving forwards: missing visits (low at present), monthly follow-up (missed positives) – currently doing sensitivity analysis stopping follow-up the first time someone hasn't had a swab in the last 14 days and results similar

Official reported estimates of the rate of new COVID-19 infections in the community in England.

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

Incidence rate per 10,000 people per day

Incidence rate per 10,000 people per day

After 17 August, the estimate has a greater level of uncertainty because of lab results still being processed

Source: Office for National Statistics

Analysis by population characteristics

- Published monthly
- Comparisons by: ethnicity; sex; age; household size; contact with confirmed or suspected cases of Covid-19; working location; health care roles; symptoms

- See https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsinthecommunityinengland/characteristicsofpeopletestingpositiveforcovid19inenglandaugust2020

Comments and next steps

- Operational delivery is a big challenge
  - Field data collection
  - Data capture and data architecture

- Estimation is also challenging because of very low positive results: started using trend estimation with MRP

- Considering adding a spatial element to the models

- Phase 3 of study: much larger unequal probability samples for whole UK

- More detailed geographical breakdown: spotting hotspots

- Will consider combining contact & trace data with household survey data: need to use a sound method