

Mail panda@pandata.org

Web pandata.org

Facebook [@pandata19](https://www.facebook.com/pandata19)

Twitter [@pandata19](https://twitter.com/pandata19)

ARTICLES

The PHE data that goes against the narrative - Hunting down symptomatic COVID-19

Aleks Nowak and
Dr Clare Craig,
with assistance from
Joel Smalley and
Dr Jonathan Engler

DECEMBER 2020

Public Health England's (PHE) [notifications of infectious diseases](#) (NOIDs) data had only 87 COVID cases listed for the week beginning 9th November compared with 145,129 cases reported on the [government dashboard](#) using another PHE data source. This is a gargantuan discrepancy and requires further investigation.

COVID-19 became a notifiable disease in March this year. Since then there has been a statutory requirement that any Registered Medical Practitioner having "reasonable grounds for suspecting" that their patient has COVID-19, has to notify their local council or Health Protection Team. The legal requirement is to notify prior to any confirmatory analysis (e.g. PCR testing). PHE collects and aggregates this data, publishing it in the notifiable disease weekly reports.

The crucially significant aspect of the NOIDs data is that it is a measure of actual disease, which by definition requires symptoms. The bulk of data offered by the government dashboard in contrast reflects positive PCR test results with no reference to symptoms. A disease requires symptomatic diagnosis which may be confirmed by testing for causative agents. The disease here being COVID-19 and the causative agent being the virus SARS-COV-2. A positive causative agent test alone does not determine disease.

I totalled the COVID-19 notifiable disease reports for COVID-19 since the start of the year; 17099 up to 15 November 2020 (across England and Wales). Dumbfounded and unable to form a thought, after forty minutes pacing the room, I reached out to Abir Ballan and Nick Hudson at PanData.org and they connected me with Dr Clare Craig to make sense of the notifiable disease dataset.



How can this vast numerical discrepancy be explained?

Doctors know they have a legal requirement to report certain specified diseases and have to do so for a large number of notifiable diseases. For some diseases, e.g. tuberculosis, the patient will have a number of interactions with doctors on the frontline, doctors making the diagnosis in the laboratory and TB nurses, each of whom may do the required reporting to the public health authorities. In this way a safety net is built in that is likely to ensure that all or nearly all cases are notified.

Some possible explanations for the now number of notified COVID-19 cases might be:

- Perhaps doctors treating COVID-19 patients have been run ragged and it is not reasonable to expect them to find the time for such administrative tasks?
- Could it be that the doctors caring for patients with COVID-19 thought that someone else would do the reporting?
- Could it be that as PHE were responsible for the testing they felt there was no need to notify them?

All three may be the case but, even if we assume the NOIDs data represents only a proportion of actual cases of COVID-19 disease, as opposed to cases of a positive detection of SARS-COV-2, there are still some noteworthy features of this data.

There were two peaks of notified COVID in England. The first peak occurred in May and the second peak began at the beginning of September, reached a maximum at the beginning of October and has returned to the summer baseline since 17 October (Figure 1).

The number of NOIDs cases are shown against the number of positive PCR test results below, and in order to better observe the relationship between the trends observed we have included a further graph in which the 2 measures are rescaled.

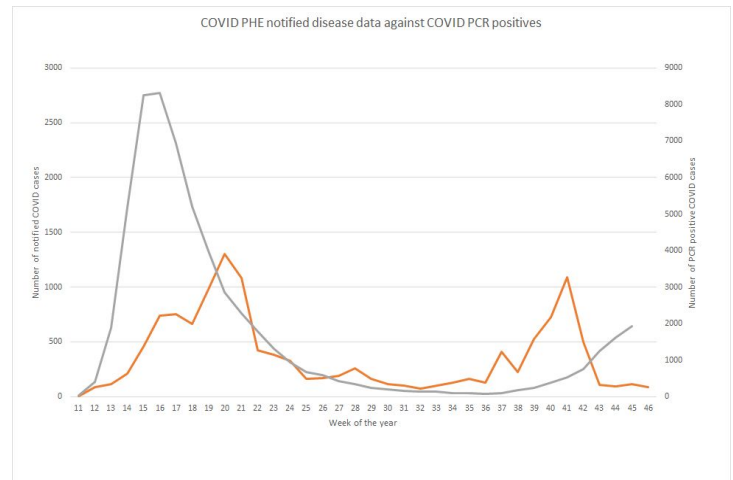
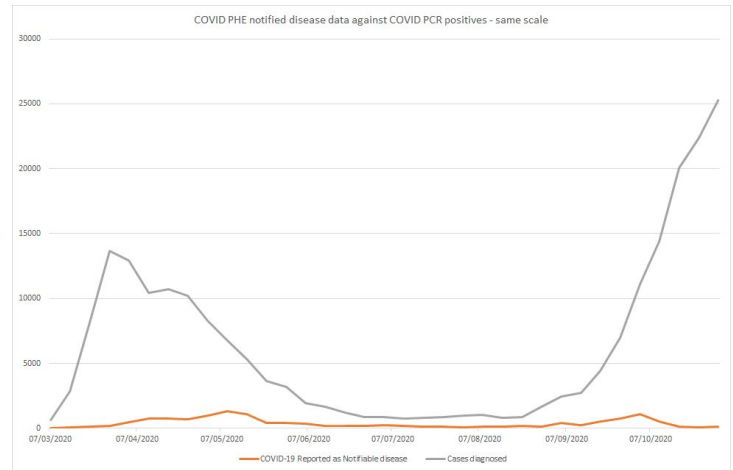


Figure 1 Notified COVID cases (orange) vs PCR positive COVID cases (grey). Top graph on same scale second graph scaled to show trends.

Viewing the same data as a percentage of cases reported demonstrates that what doctors were notifying as symptomatic COVID via NOIDs became an insignificant proportion of the total PCR positive COVID cases in mid-October (Figure 2).

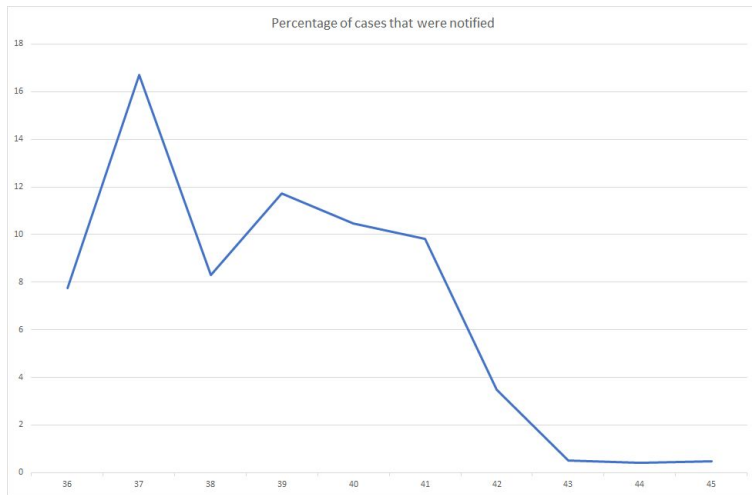
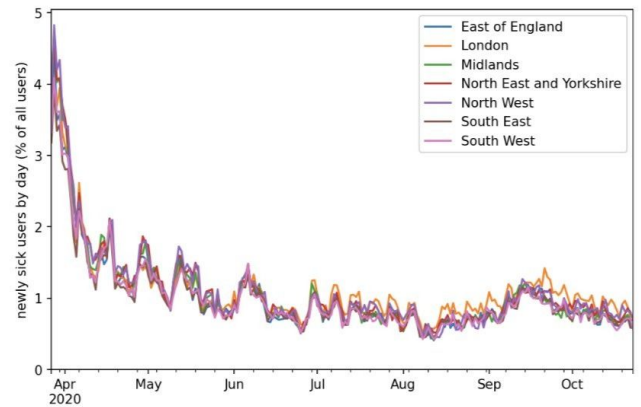


Figure 2 Percentage of PCR positive COVID cases that were notified to PHE via NOIDs

The significance of a September rise in notified cases that returned to normal by October is that other data shows a similar pattern. The Zoe App symptom tracker has a large number of nationwide participants who enter their symptoms and test data to track COVID symptomatology. This data also showed a September rise and mid-September peak before returning to a previous baseline (figure 3).

Data from NHS triage shows the numbers of people phoning 111 or using NHS online which the system categorise as “COVID like” (figure 4). These two data sources also show a mid-September spike before returning to a baseline above the summer baseline.

Newly Sick Users by Day in NHS England Regions (% of Active Users)



Positive Test Results by Day (% of Invited Test Takers)

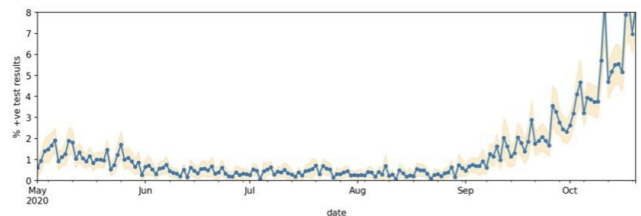


Figure 3 Contains two charts, the first plots ZoeApp symptom tracker data showing a mid-September peak in symptoms in the community which then returns to baseline while the second chart shows the positive test results continue to climb while symptoms decline. Graph from @timspector.

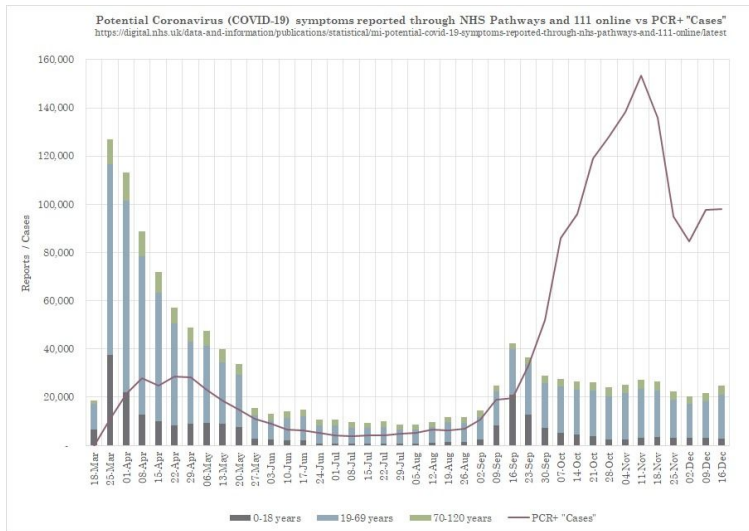


Figure 4 Solid coloured bars show *daily* numbers of patients triaged by NHS 111 and categorised as COVID like (by age) with a peak on 15th September. The red line shows the **weekly** 'cases' defined by a single positive PCR test result. Scales are different but demonstrate trends. Graph by @realjoelsmalley.

There are therefore three datasets all showing a second wave that begins in September, peaks two weeks later and returns to baseline by the end of September:

1. PHE Notifiable Infectious Disease
2. Community symptom tracking
3. NHS triage classification of patients

All three datasets concur but contradict the PCR positive test results.

There are two other diagnostic tests for COVID:

1. Lateral flow tests which test for viral proteins that form part of the viral particles themselves
2. Antibody testing which confirm infection two to five weeks afterwards

Results of lateral flow testing in Liverpool and Merthyr Tydfil show a consistent positive rate of <0.8%. This is higher than the expected false positive rate calculated with clear cut positive and negative cases but it is often the case that the false positive rate is higher when testing with real world ambiguity. It is striking that the positive rate for this test is the same over time and in different places. Lateral flow tests will not diagnose every case. In fact, 20% will be missed which is the same false negative rate as seen with most PCR testing. This is insufficient to account for the discrepancy between ONS PCR testing and lateral flow testing of the population at random. The ONS predicted 2.3% of Liverpool had COVID on 11th November but lateral flow testing found only a third of that - 0.7%. Even if 20% of cases had been missed by lateral flow tests, that is not enough to bridge the gap.

Antibody testing of the population has failed to show a rise in levels despite plenty of time since the rise in cases.

Figure 51: 4-weekly rolling SARS-CoV-2 antibody seroprevalence (% seropositive) in blood donors by region, using Euroimmun test; error bars show 95% confidence intervals

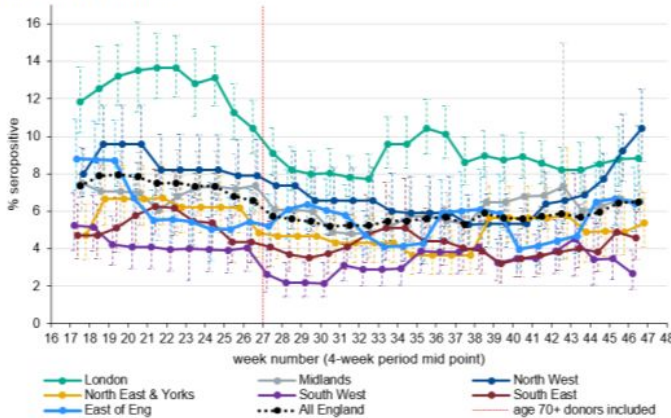
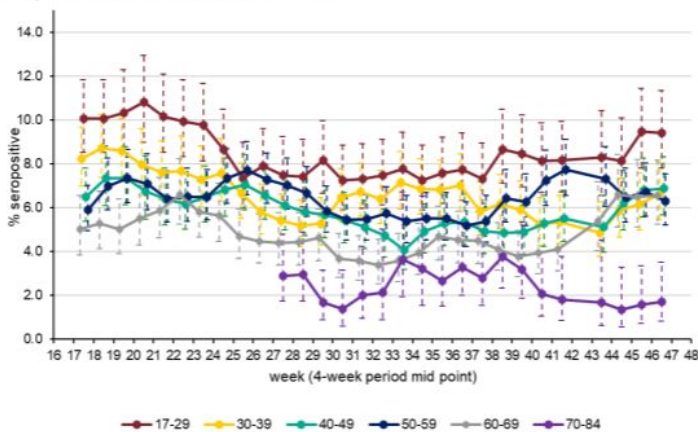


Figure 52: Population weighted 4-weekly rolling SARS-CoV-2 antibody seroprevalence (% seropositive) in blood donors by age group, using Euroimmun test; error bars show 95% confidence intervals



Figures from [Public Health England Weekly national Influenza and COVID19 surveillance report](#)

Again two sources of data, lateral flow testing and antibody testing concur and contradict results from PCR testing.

Could it be that reports of Notifiable Infectious Diseases - being filtered by the need for symptoms to gain the attention of GPs - are actually a more representative measure, at least in terms of trends, of “real” COVID-19 than PCR testing?

It cannot have gone unnoticed by PHE that this notifiable disease data is very divergent from published case data. The Government must urgently review their processes:

1. Only diagnose disease in symptomatic patients
2. Hospital diagnosis should be based on alternative bedside testing for rapid and reliable results (E.g. Lateral flow tests)
3. PCR testing must be halted until the quality of the results has been properly audited

ABOUT THE AUTHORS



Aleks Nowak is a Blockchain Technology Entrepreneur. Twitter: @an648



Dr Clare Craig has been a pathologist since 2001 for the NHS in the UK and more recently as the day to day pathology lead for the cancer arm of the 100,000 Genomes Project and working for an AI startup in cancer diagnostics. Twitter: @ClareCraigPath



Joel Smalley is a self-taught empiricist with an MBA from University of Toronto. Data science, econometrics, quantitative analysis. Early career trading equity and credit derivatives then quant, specialising in time series and stochastic analysis. Twitter: @RealJoeSmalley



Dr Jonathan Engler is qualified in medicine and law, but has spent most of his career in the pharma industry. He founded a company using tech to coordinate clinical trials, and after selling it, currently runs a small biotech doing reasonably clever things with cancer stem-cells. Twitter: @jengleruk

REFERENCES

1. <https://www.gov.uk/government/publications/notifiable-diseases-weekly-reports-for-2020>
2. <https://coronavirus.data.gov.uk/details/cases>
3. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/942969/Weekly_Flu_and_COVID-19_report_w50_FINAL.PDF

Photo by [Scott Graham](#) on [Unsplash](#)