

Coronavirus: Data can save lives, data can cost lives - and this latest testing blunder will likely prove it

 [news.sky.com/story/coronavirus-data-can-save-lives-data-can-cost-lives-and-this-latest-testing-blunder-will-likely-prove-it-12090904](https://www.news.sky.com/story/coronavirus-data-can-save-lives-data-can-cost-lives-and-this-latest-testing-blunder-will-likely-prove-it-12090904)



Data can save lives; data can cost lives.

That might sound a little melodramatic until you recall that it is ultimately the very point of Britain's **COVID-19** test and trace strategy.

The sooner we detect the disease and marshal that information to prevent infected people passing COVID on to their contacts, the sooner we can bring this pandemic under control and return to some semblance of normality.



Will the missing COVID cases have an impact?

Numbers are what undergird this process. Data on who has the disease, data on their contacts: if we lose control of that data then we have effectively lost control of the **coronavirus** all over again.

Which is why what happened in the past week with England's COVID-19 testing data is so disturbing. The data on **some 16,000 positive cases** of this disease were briefly lost and have only belatedly been recovered and passed on to contact tracers.

Some of those infection alerts were delayed by a few days, others by nearly a week.

In a pandemic where every hour counts, that is not far short of a disaster. It makes it highly likely that some of those contacts who were not reached in time will unknowingly have been spreading COVID-19.

It means that in some clusters, the disease was allowed to spread unchecked. It means more people will have caught it; some of those people will probably die.

And all because of an astonishingly elementary spreadsheet mistake at Public Health England (PHE).

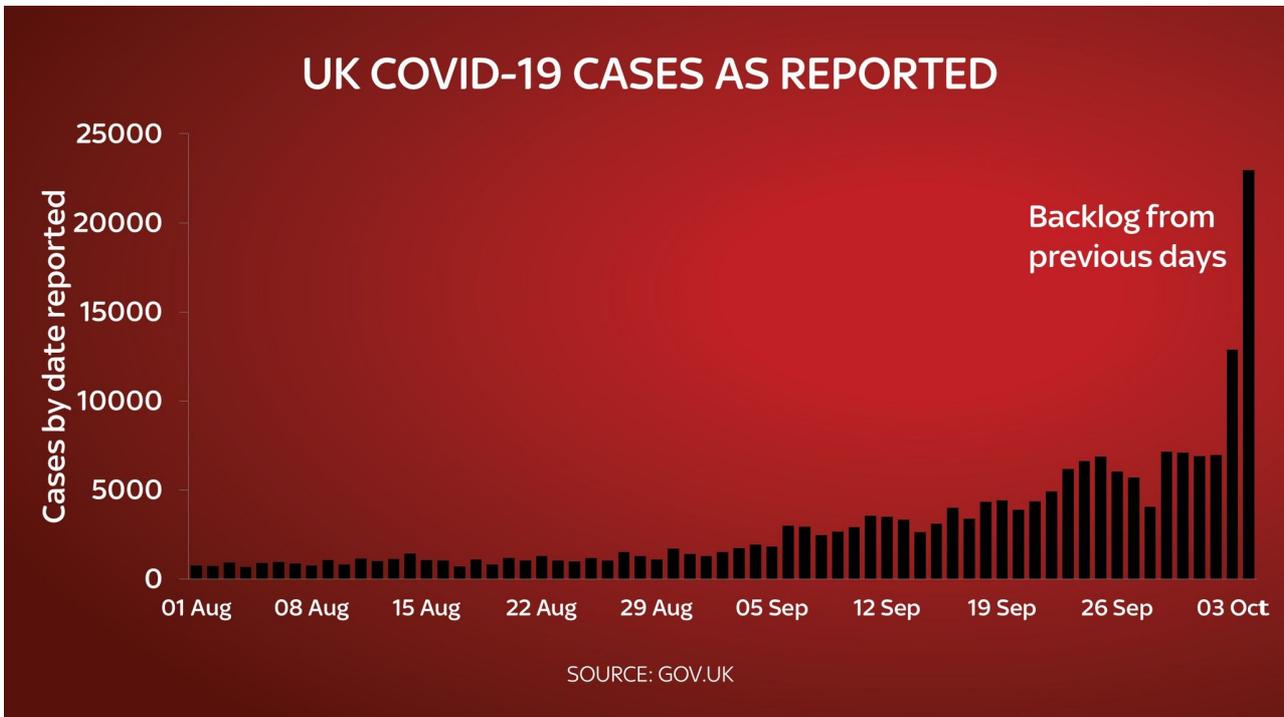


Image: UK cases as reported

Now, if you've spent any time fiddling around with data you've probably spent some time with your nose inside a spreadsheet program like Excel. And as any data nerd will tell you, Excel is both amazing and at the same time utterly frustrating.

It has all sorts of incredible functions which allow you to fiddle and query and visualise data. You can take a table of data and you can very quickly analyse it, turn it upside down, shake it around and work out what's really going on underneath the surface.

It is - and I say this as someone who spends much of his days buried inside it - a thing of wonder.

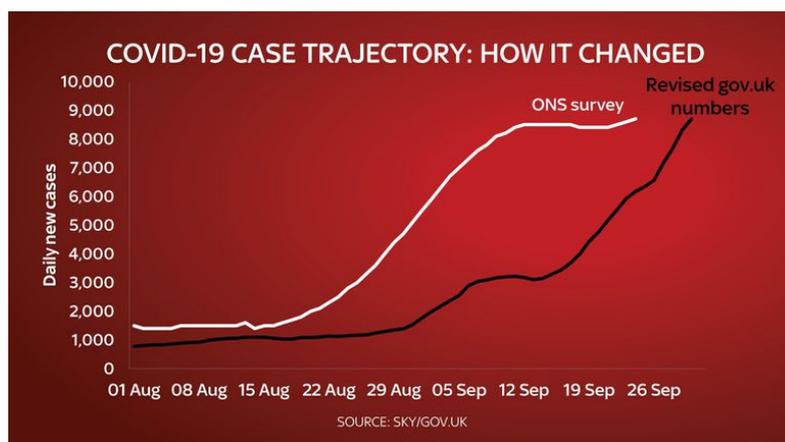


Image: The cases from the ONS data against the testing data

But like every clever tool, spreadsheet programs like Excel also have a whole host of serious limitations. For instance, up until the most recent versions, there was a somewhat arbitrary limit on the number of rows of data you could import into the programme.

Anything above 65,536 rows and it simply won't work. The latest version of the software raised that limit to a million rows, but many computers - especially the ones inside government offices - are still running the older version.

Anyone working with big databases is well aware of those limitations. It is why they rarely trust Excel for managing big databases, using other dedicated database services instead.

Indeed, PHE has a pretty robust database it has used for years to collate national test results for various diseases.

The problem is while the Second Generation Surveillance System (SGSS), as it's called, is plugged into the existing pathology labs that pre-date COVID-19, for some reason no-one has yet managed to connect it to the new labs which have been created during the course of this pandemic.



Image: The current trajectory mapped against the government warnings two weeks ago

The friction between these two systems - Pillar 1 labs, which constitute the established labs in hospitals around the country, and Pillar 2, which constitute the new centralised mostly privately-run labs created specifically in the face of the disease - is one of the main problems **which has bedevilled Britain's testing system**.

This latest data disaster is only the latest episode.

Rather than feeding its results securely straight into SGSS, as Pillar 1 pathology labs do, the test results from Pillar 2 labs arrive at PHE in the form of a CSV file.

CSV, in case you haven't yet encountered it, is about the most basic spreadsheet format that exists, with data separated by commas.

That CSV file is then automatically fed into an Excel template, which then feeds it into the government's testing dashboard.

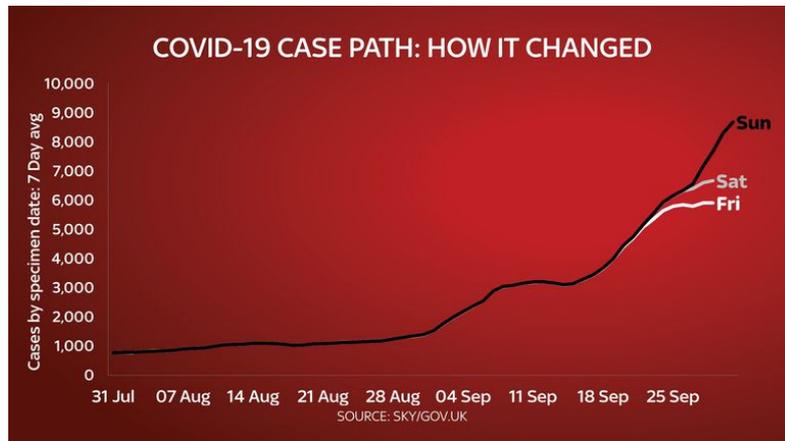


Image: The jump in cases after the backlog was added in to the data

In data management terms, this is a little like putting together a car with sellotape, for reasons PHE discovered on Friday.

While all the other indicators were suggesting the case numbers in the UK were rising pretty rapidly, the numbers being displayed on the government's dashboard seemed to be stalling, and then falling.

The technicians at PHE opened up the computer doing the Excel conversion and discovered something alarming: it hadn't included all the data from those Pillar 2 labs.

It turns out that as the number of cases mounted the number of rows on the spreadsheet was getting longer and longer and suddenly PHE's version of Excel - which is thought to be the older version - came up against that 65,536 row limit.



PM says current COVID numbers are 'realistic'

Thousands of rows of data - which is to say information on cases - were simply left out.

There are many unsettling things about this but perhaps the most unsettling is that this process - with data sellotaped together - is at the very apex of Britain's COVID-19 management system.

For only after PHE has processed the data is it passed onto the contact tracers who can then get to work trying to isolate those who have been in contact with infected people.

In the event, about 16,000 cases were not processed immediately and were only passed onto contact tracers on Saturday.

According to PHE officials, around 12,000 of them were delayed by only one to three days - but the remaining 4,000 were delayed by as much as a week.

Breakdown of missing cases

	Missing cases	Original figure	Actual total
25 September	957	6,874	7,831
26 September	744	6,042	6,786
27 September	757	5,693	6,450
28 September	0	4,044	4,044
29 September	1,415	7,143	8,558
30 September	3,049	7,108	10,157
01 October	4,133	6,914	11,047
02 October	4,786	6,968	11,754

According to insiders, the spreadsheet problem has now been addressed - which is to say they are chopping the incoming CSV files into smaller sizes to allow them to fit into Excel.

At this stage it is perhaps worth underlining that around £12bn has been budgeted for the test and trace system - more than almost any other government investment programme in modern history.

It is double the amount the UK is spending on its two aircraft carriers and equivalent to almost £450 for every household in this country.

Many will ask how so much could have been set aside for this scheme, yet it has been undone by a known issue in a computer program which could have been solved by spending about £100 on an upgrade.

The other consequence of the data revisions is to change our impressions of the spread of the disease. For the national dashboard on which those data are displayed now suddenly looks very different to how it did on Friday.

Here the implications are somewhat less serious, if only because no-one much trusted the daily case figures anyway.

Still: on the basis of this data series which most of us have been treating with a little caution, it does indeed change the picture a bit.

Up until the revisions over the weekend the average daily increase in positive COVID-19 tests (by date of test) was running at just under 5% over the past fortnight.

As of this new data it's running at around 7%.

That might not sound like much of a difference - and be assured it's still a long way shy of the 10% plus growth rate (the equivalent of cases doubling every seven days) Sir Patrick Vallance warned of a couple of weeks ago.

Even so, it makes for uncomfortable reading.

It means, for instance, that when you compare the UK to the French and Spanish trajectories for the disease, the UK goes from being below their lines - in other words having a less severe outbreak - to being above their lines.

'We can't change the past': Coffey on 'IT failure'

But most people pay more attention to other measures such as the Office for National Statistics survey of the infection and hospital admissions.

And on the basis of these measures the picture remains as it did before the weekend: the disease is spreading but the spread looks less rapid than it did a few weeks ago.

Even so, the great worry about this episode is not only that it harms faith in the system designed to protect us from the disease; it has already allowed more people to be infected.

And all because no-one paid enough attention to the data.