

Botched Excel import may have caused loss of 15,841 UK COVID-19 cases

Ars Technica

reader comments

191 with 114 posters participating, including story author

Share this story

Public Health England admitted on Sunday that the agency has under-reported COVID-19 infections by 15,841 cases in recent days due to a “technical issue.” The missing positive tests were conducted between September 25 and October 2 and have since been added to national statistics, the agency said.

PHE didn't explain the nature of the technical issue, but a number of British news sources have pointed the finger at Microsoft Excel. Here's how the Guardian describes the issue:

PHE was responsible for collating the test results from public and private labs, and publishing the daily updates on case count and tests performed.

In this case, the Guardian understands, one lab had sent its daily test report to PHE in the form of a CSV file – the simplest possible database format, just a list of values separated by commas. That report was then loaded into Microsoft Excel, and the new tests at the bottom were added to the main database.

But while CSV files can be any size, Microsoft Excel files

can only be 1,048,576 rows long. When a CSV file longer than that is opened, the bottom rows get cut off and are no longer displayed. That means that, once the lab had performed more than a million tests, it was only a matter of time before its reports failed to be read by PHE.

The agency says it will take precautions to make sure an error like this doesn't happen in the future.

In addition to understating the number of COVID cases in official statistics, the glitch also hampered contact tracing efforts because data about the missing positive tests were not passed along to contact tracers. Officials stressed that the test subjects themselves got their test results as normal.

The United Kingdom is currently in the middle of a second surge in coronavirus infections. Infections in the country peaked in April and then declined to a few hundred per day in July. But cases have soared over the last month, and in recent days the UK has reported more than 10,000 new cases per day (including the previously missing cases). That's about as many as the US on a per capita basis.

Update (1:25pm): The BBC offers a different Excel-related explanation for the error:

PHE had set up an automatic process to pull this data together into Excel templates so that it could then be uploaded to a central system and made available to the NHS Test and Trace team as well as other government computer dashboards.

The problem is that the PHE developers picked an old file format to do this – known as XLS.

As a consequence, each template could handle only about 65,000 rows of data rather than the one million-plus rows that Excel is actually capable of.

And since each test result created several rows of data, in practice it meant that each template was limited to about 1,400 cases. When that total was reached, further cases were simply left off.

<https://arstechnica.com/?p=1711742>