

## ARTICLES

### A miscarriage of diagnosis

Dr Clare Craig

DECEMBER 2020

PCR testing for COVID-19 aims to detect individuals with a high likelihood of being infectious. However, false positive test results lead to false diagnoses, unnecessary measures and distort the overall picture of the pandemic.

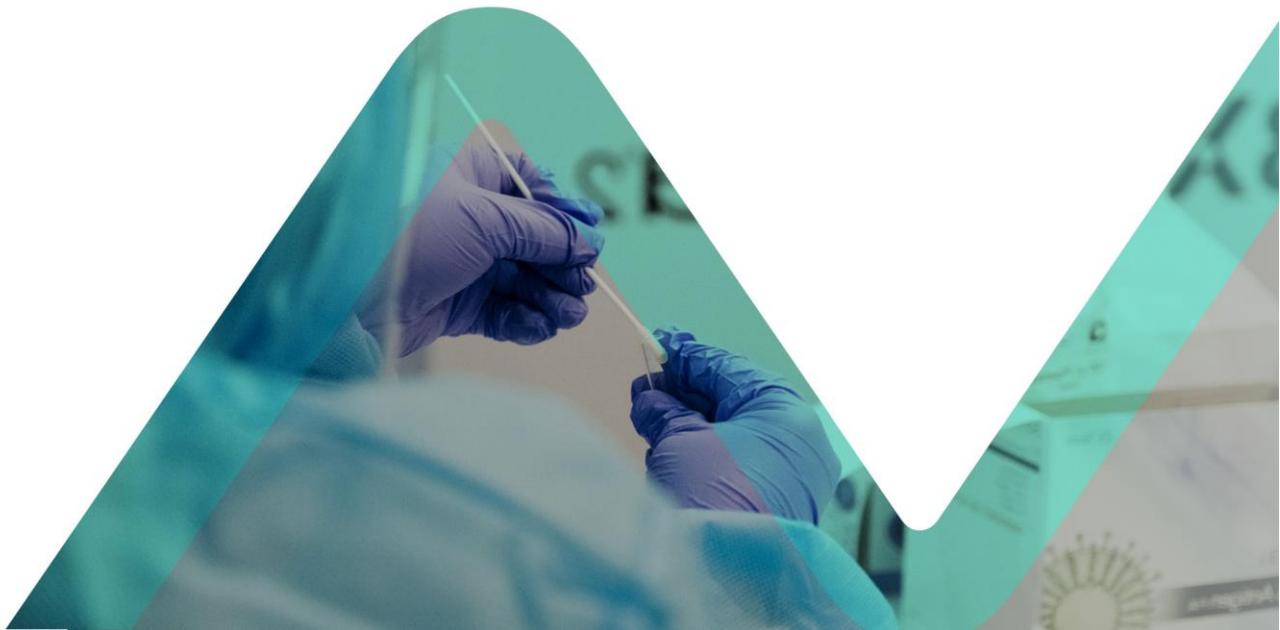
The PCR test false positive results have more than one cause and productive conversations about them require the five categories discussed here to be distinguished from one another.

The **operational false positive rate** refers to the rate of error across the whole process. This will vary from day to day so the rate should be measured as a tendency to a mean, not taken as the minimum. Each laboratory will have its own operational false positive rate and this can change over time depending on the factors below.

#### 1. Profiling errors

The person being tested has a significant bearing on the false positive rate. Any positive pregnancy test from, say, testing children in a reception class at primary school must be a false positive. Likewise testing asymptomatic people (eg. airport testing) for COVID-19 is much more likely to produce false positive results than testing symptomatic patients.

For unknown reasons, some groups within communities can have a higher baseline false positive rate. This is a frequent problem we see, for example, in breast and cervical cancer screening in young women – a reason why those programmes do not screen young women. For COVID-19 a similar unexpected level of false positives was seen in the





summer with people in their 20s. When this subpopulation with a high false positive rate was discovered, they were targeted for more testing. We now know they were false positive results because the evidence from spring across the world proves that genuine COVID-19 outbreaks spread rapidly between age groups. This did not happen throughout August, which proves that the “outbreak” amongst young people was a pseudo-epidemic made up of false positives.

It is important to target testing at people who have symptoms and provide a high clinical suspicion of the disease being tested for. This targeting of a subpopulation because they have a high false positive rate is bad profiling. The more people in this subpopulation you target, the higher the false positive rate will be for those tested as a whole.

## **2. Mistaken identity**

The likely underlying cause of the false positives in young people was mistaken identity. When testing for RNA (the viral equivalent of DNA used for replication), the test should be able to distinguish between sequences that are unique to COVID-19 and sequences seen on other viruses, or even in human DNA. However, no test is perfect.

Human DNA [has been mistaken](#) for a different coronavirus when doing PCR testing. The human genome comprises three billion letters of code. While none of it may be an exact match for what the PCR test should be detecting, a near match could result in errors in a proportion of the tests. This type of mistaken identity could lead to particular subpopulations being targeted for testing, creating profiling errors.

A 2003 outbreak of SARS-1 [in a care home in British Columbia](#), turned out to be a common cold causing coronavirus. Coronaviruses are a family of viruses and, although the spike protein of the COVID-19 virus is unique, the rest of the virus has many features similar to other common colds. These similarities can cause mistakes in PCR testing. As coronaviruses are seasonal, this type of mistaken identity can cause a seasonal variation in the false positive rate.

## **3. Contamination of the chain of evidence**

There is a chain of evidence from the sample being taken, through delivery to the laboratory, checking in of samples and then opening and working on them. Contamination can happen at any stage. This contamination may come from the individuals carrying out the work or from other patients' samples once in the laboratory.

Claims that PPE would be effective at preventing contamination from swab takers etc. is like claiming that wearing chain mail would prevent you from getting sandy on a beach. A delivery driver who is post-infective and shedding RNA could contaminate the containers the samples are transported in. Whoever opens those containers could then transfer the RNA to the contents. If the same gloves are worn when opening numerous patient sample pots, then the possibility for contamination between samples will be high. Many readers may have seen the disturbing images of an undercover Dispatches reporter showing how some samples have been handled when they arrive at a lab.

Contamination is an issue largely because of the nature of the test rather than sloppy handling. Having turned the RNA into DNA, the second step in testing is to multiply the DNA by one billion to a trillion times. That means that even with highly competent



sample handling, the risk of contamination will remain because only the tiniest fragment of contaminant RNA can create a false positive test result. Reducing the number of times the DNA is multiplied reduces the chance of these errors, but even then, not to zero.

The risk of cross contamination from true positives samples will be greatest when there is real COVID amongst those samples being tested.

#### **4. Equipment errors**

The testing equipment itself will have a low and fairly constant false positive rate. This is of the least significance but has had the most effort put into understanding it. It is possible to calculate based on retesting samples with different test kits. There seems to be a general misunderstanding that this is the only cause of false positive error and that because it is a low value there is no false positive problem.

#### **5. Burden of proof**

As well as choosing a reasonable cycle threshold to reduce contamination errors, other variations in the criteria used to determine positivity will lead to differences in the false positive rate. It is standard practice to test for three genes belonging to the COVID-19 virus. However, if positive is defined as the presence of a single gene rather than all three then the false positive rate will be higher.

For example, the REACT study at Imperial carried out calibration between PCR tests in commercial laboratories and the same samples tested in Public Health England laboratories. They found a 57% false positive rate in May. To minimise this error, they used different criteria to the commercial laboratories. Instead of reporting on one gene at any threshold, they chose to define the presence of one gene below a cycle threshold of 37 or the presence of two genes, as positive.

These five types of false positive errors make up the operational false positive rate. Changes in who is targeted, seasonal infections and laboratory quality standards can lead to changes in the false positive rate over time. The five types of false positives will vary between laboratories so investigations as to the rate at one laboratory cannot be extrapolated to another, and each has its own interaction with underlying community prevalence rates. Therefore, the overall epidemiological false positive rate will vary by place, time and testing strategy.

## ABOUT THE AUTHOR



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

## REFERENCES

1. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7128111/>
2. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2095096/>

Photo by [JC Gellidon](#) on [Unsplash](#)