Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
JUST how many people have been infected with the coronavirus? Statistics are trickling in from cities and countries around the world, but the figures vary hugely. Some regions are reporting that less than 1 per cent of people have been infected, and others that over half the population has had the virus. How are these figures calculated, and which can we trust? Determining the true prevalence of coronavirus infection will be important for understanding how the virus spreads and limiting its damage.

The reporting of coronavirus cases varies drastically around the world. Tim Russell and his colleagues at the London School of Hygiene and Tropical Medicine have estimated that, as of 15 June, more than 95 per cent of symptomatic cases have been reported in some countries, including Ghana, Kazakhstan, Morocco and Oman.

However, the team estimates that only 35 per cent of symptomatic cases have been reported in some countries, including the UK, which has an estimated reporting rate of about 19 per cent. Sweden is estimated to have reported only 14 per cent, and Yemen just 5 per cent.

What these statistics don’t reflect is the number of symptomless cases, which some evidence suggests can account for between a quarter and half of all coronavirus infections.

Two ways to estimate

When it comes to estimating prevalence, there have been two main approaches. Researchers have either tested a sample of people in a population and directly reported those numbers, or predicted how the virus has affected a population using mathematical models.

As a starting point, many modellers have turned to the case of the Diamond Princess, a cruise ship that experienced one of the first large outbreaks of the coronavirus. Most of the 3711 people on board were tested for the virus – 712 were found to have been infected, and at least 13 people have died.

Russell and his colleagues have used this and other data sources, such as European Centre for Disease Prevention and Control data, to predict how many people have been infected with coronavirus in various countries. Because the quality and quantity of the testing and reporting is inconsistent across the world, the team says it is difficult to get an accurate picture of the true number of infections.

“The quality and quantity of testing is inconsistent, so it is difficult to get a true number of those infected”

It may be possible to come up with an estimated global rate of infection, but due to missing data, it may not be a very good one. One way to get around this is to look at the overall number of deaths from all causes. “If you’re seeing way more deaths, you’re probably not detecting anywhere near as many cases as you have,” says Russell.

Henrik Salje at the University of Cambridge and his colleagues have done similar modelling for France. Their technique involves incorporating data from sources like the Diamond Princess and hospital records, and filling in gaps by estimating how many people might have had infections that aren’t severe enough to require hospitalisation, which is more common for women and younger individuals.

By their estimate, as of 11 May, around 4.4 per cent of the population of France had been infected with the coronavirus. The figure varies by region, and the hardest hit area of the Ile-de-France region, which includes Paris, is thought to have had a prevalence of 9.9 per cent by then.

These figures are broadly in line with others resulting from direct antibody tests, which can reveal whether a person’s immune response has encountered the coronavirus in the past.

So far, antibody tests in Spain, Italy and England put the proportion of people infected at between 2 and 5 per cent, although, again, the prevalence
is higher for densely populated cities like London and Madrid.

These tests work by searching for specific antibodies – proteins made by the immune system that help defend against infection. In the case of the new coronavirus, most antibody tests search for immunoglobulin G, which is the most common type of antibody and typically sticks around in the blood longer than others, says Juliet Bryant, a consultant for the World Health Organization.

However, we don’t yet know the role such antibodies might play in conferring any immunity to the coronavirus.

A global effort

In May, Niklas Bobrovitz at the University of Toronto in Canada and his colleagues formed a team of scientists in six countries to collate and assess blood test data to determine the proportion of people globally who have been infected with the coronavirus.

Many regional figures come from studies that haven’t been through the peer-review process yet, and some come from public health bodies or private health agencies. Sometimes, the information comes in the form of a tweet from an elected official.

So far, there are more than 100 of these prevalence figures, says Bobrovitz, although that number is growing on a daily basis. “There’s a huge variation, between 0.4 and 59 per cent,” he says.

The team can’t be sure what is driving these differences. There could be real differences in the prevalence rate, perhaps due to the age or health of the population, or policies to control transmission of the virus. “Very likely, it’s down to differences in the study methodology used,” says Bobrovitz.

There are a variety of coronavirus tests, and each will have its own sensitivity and specificity. These two measures give an idea of how many test results will be false negatives or false positives, respectively. In some cases, studies don’t report the accuracy of their tests.

While some researchers have looked at antibody levels in the general population, many have focused on groups that are easier to study, such as healthcare workers and blood donors. A large majority of the studies that Bobrovitz and his colleagues have assessed so far are considered to have a moderate or high risk of bias in their reporting. A study that includes blood donors, for example, is likely to over-represent young people.

26% of people have been infected with coronavirus in Stockholm

One of the more robust studies has been conducted in Spain, where authorities recruited more than 60,000 members of the general population. So far, the results of testing suggest that about 5 per cent of the population across Spain has been infected, although the figure is 10 per cent in Madrid.

Far from herd immunity

Despite the problems with these studies, it is clear that even countries with high prevalence are nowhere near the level needed to allow herd immunity, which can stop infections spreading through populations. For the majority of known viruses, at least 70 per cent of the population must have immunity to achieve this communal protection.

This level of infection hasn’t been observed in any study, not even in Sweden. The country has become a poster child for the anti-lockdown movement, due to the comparative lack of restrictions on daily life to control transmission of the virus. Sweden has reported more deaths per capita than neighbouring countries, but modelling by the country’s health authority estimates that only around 26 per cent of the population of Stockholm County had been infected with the virus by the start of May.

Salje thinks this is probably an overestimate. The virus is likely to cause a similar percentage of deaths across countries, he says. If more than a quarter of the Swedish population had been infected, “you would expect to have seen many more deaths”, he says.

Getting our heads around how many people have been infected by the coronavirus isn’t simply a matter of tracking the progress of the pandemic. Understanding how many people have had a virus can help shape public health policies by showing what works and what doesn’t when it comes to limiting the spread, says Bryant.

Working out overall numbers and whether people have had the virus would therefore be of great benefit. But even if an antibody test were perfect, and delivered no false positive or negative results, we still wouldn’t know if previously infected people will have immunity against the coronavirus or how long such immunity might last.

And when tests return false positive and negative results, they can give people the wrong idea about whether they are likely to catch or spread the virus. However, when we know the rate of false results, we can still get a good idea at the population level of how many people have had the virus. “Even a crappy test can be useful if you use it at the population level,” says Bryant.