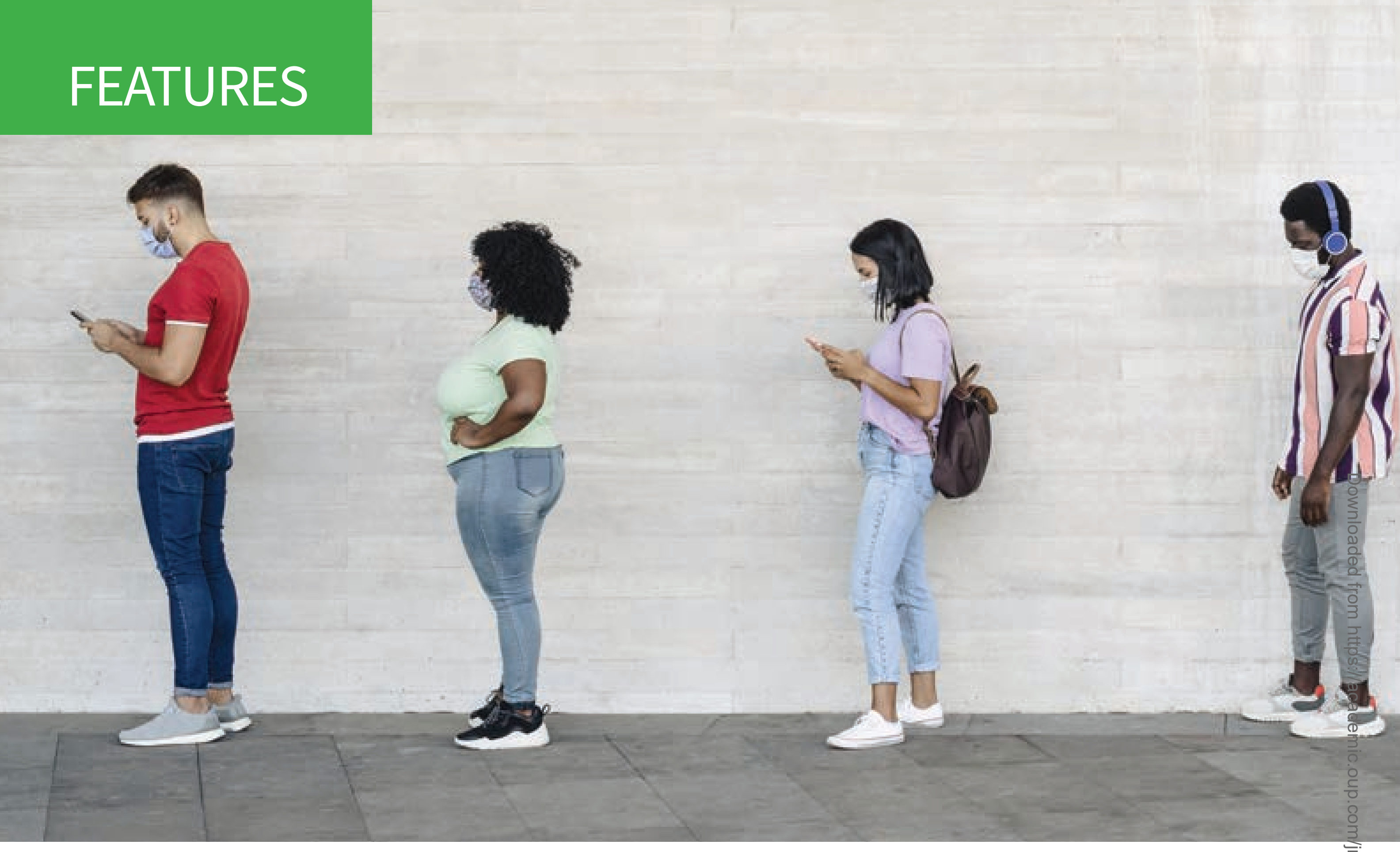

The story of the R number

How an obscure epidemiological figure took over our lives

Part 1: History

The R number was the "breakout star" of the early weeks of the Covid-19 lockdowns, swiftly taking up residence in the general public's consciousness and conversations. In this new six-part series, **Gavin Freeguard** will delve into the evolution of the R number and how it was modelled, used by the UK government and the media and regarded by the average person during the pandemic. He will also consider its limitations, lessons learned and how such statistical tools should be managed for disease planning and response in the future. In this first instalment, he introduces the history of this statistical concept.

Alessandro Biascioli/iStockPhoto.com

Covid-19 did not only dominate our lives in April 2020. It also dominated the list of new words entered into the *Oxford English Dictionary*.

Alongside Covid-19 itself (noun, "An acute respiratory illness in humans caused by a coronavirus"), the vocabulary of the virus included "self-quarantine", "social distancing", "infodemic", "flatten the curve", "personal protective equipment", "elbow bump", "WFH" and much else. But nestled among this pantheon of new pandemic words was a number, one that would shape our conversations, our politics, our lives for the next 18 months like no other: "Basic reproduction number (R_0): The average number of cases of an infectious disease arising by transmission from a single infected individual, in a population that has not previously encountered the disease."

"There have been many important figures in this pandemic," wrote *The Times* in January 2021, "but one has come to tower over the rest: the reproduction rate. The R number, as everyone calls it, has been used by the government to justify imposing and lifting lockdowns. Indeed while there are many important numbers — gross domestic product, parliamentary majorities, interest rates — few can compete right now with R" (tinyurl.com/v7j6cth9).

Descriptions of it at the start of the pandemic made R the star of the disaster movie reality we lived through. And it wasn't just a breakout star of the UK's coronavirus press conferences; in Germany, (then) Chancellor Angela Merkel made the most of her scientific background to explain the meaning of R and its consequences to the public (tinyurl.com/mva7urw5).

But for others, the "obsession" (Professor Linda Bauld, University of Edinburgh) with "the pandemic's misunderstood metric" (*Nature*: tinyurl.com/y3sr6n6m) has been "a distraction", an "unhelpful focus"; as the University of Edinburgh's Professor Mark Woolhouse told one parliamentary select committee, "we've created a monster".

How did this epidemiological number come to dominate our discourse? How useful is it? And where does it come from?

The history of R

> It is sometimes interesting to stand back a bit and view diseases, not as we are always exhorted to do in medicine, in the way they affect individuals of varying make-up, but as entities in themselves with a natural life history, and to consider what factors affect that life history.
>
> George Macdonald[1]

The story of R begins not with Covid-19 but with a very different disease: malaria.

We need to step back a couple of generations to understand the transmission of the idea for R. Sir Ronald Ross won the 1902 Nobel Prize for his work on the transmission of malaria – he later became head of the Ross Institute, now the Department of Tropical Hygiene at the London School of Hygiene and Tropical

Gavin Freeguard is a freelance consultant specialising in data, an associate at the Institute for Government, policy associate at Connected by Data and special adviser at the Open Data Institute. He was originally commissioned by Understanding Patient Data to develop these articles.

Medicine (LSHTM). Ross became interested in mosquitoes while stationed as a surgeon in Bangalore, India – removing a water tank had stopped them bothering him. On sabbatical in London, he met Sir Francis Manson, whose studies in China had shown mosquitoes carried filariae, microscopic worms that spread disease. They wondered if mosquitoes might also carry malaria. Ross proved malaria was spread by mosquito bites.

He thought that if the number of recoveries from malaria outnumbered new infections, the disease would struggle to survive. He proved this first in the field, in Sierra Leone and Egypt, by disrupting mosquito populations and finding that malaria infections fell. Then, he turned to maths, in his 1910 book, *The Prevention of Malaria*.[2] According to current LSHTM professor, Adam Kucharski, in his rather more recent book, *The Rules of Contagion*,[3] Ross outlined two different approaches to studying disease: descriptive, "starting with real life data and working backwards to identify predictable patterns"; and mechanistic, "outlining the main processes that influenced transmission" and building conceptual, mathematical models. Ross favoured the latter. He considered epidemiology "a mathematical subject", and that "fewer absurd mistakes would be made" if it were regarded as such.

Ross worked further on disease dynamics with mathematician Hilda Hudson (limited by her working on aircraft engineering for the Air Ministry during the First World War). The whole field took a mathematical turn. The Scottish scientists A. G. McKendrick and W. O. Kermack wondered what would cause a disease epidemic to end. They classified individuals in a population as either susceptible, infected or recovered, and suggested a critical threshold – an epidemic would not take hold unless rates of infection exceeded rates of recovery (and death).

The creation of R

In 1952, George Macdonald – one of Ross's successors as head of the Ross Institute – wrote a journal article, "The Analysis of Equilibrium in Malaria".[4] Stored like some dangerous sample in the depths of the Wellcome Collection in London, viewable only on request by an accredited (and, during the pandemic, mask-wearing) researcher in the bright, white antiseptic Rare Materials Room, Macdonald's article marks a key shift in our understanding and analysis of disease transmission.

It begins by reciting his own previous work, which contended that the malaria infection rates in mosquitoes and in humans must be mutually dependent, finding the equilibrium of his title. He referenced Ross on "epidemic happenings" – "some minimum number of mosquitoes, above zero, was needed to keep transmission going" – and supposed that while doubling the number of mosquitoes would not simply double the incidence of malaria, there was reason to think there would be a "progressive growth": an increase in cases would, in turn, increase the proportion of infected mosquitoes, and so on. He criticised those since Ross who had "deliberately developed" his field into "abstract mathematical study", and thought the growth in malaria could only be analysed by returning to Ross's methods, "with an effort to translate the mathematical symbols back into ordinary epidemiological terms".

In a mathematical section at the end of his article, Macdonald defines the "basic reproduction rate of malaria" as "The number of infections distributed in a community as the direct result of the presence in it of a single primary non-immune case". He would develop the idea of the "reproduction rate" – soon known as z_0 rather than the R we are now familiar with – over the next 15 years. Calculating it perfectly might be impossible – it was "only a concept and not an actual event in nature" (immunity makes a difference to transmission in the real world) – but one should at least "form a concept" of the rate, since "The object of all [disease] control is to reduce the reproduction rate below one".

In his final article, published posthumously in 1968, Macdonald wrote about how "The practicability of quantitative dynamic studies has … been greatly changed by the facility of the computer whereby previous intractable aspects can now be handled with ease".[5] The transmission and evolution of Macdonald's work continued, as others estimated reproduction rates for malaria, built "Ross–Macdonald" mathematical models, and corrected some of his errors. A couple of decades after Macdonald's article, German mathematician Klaus Dietz (who also traced "reproduction rate" back to 1880s demography, where it meant the number of females born from one female during her reproductive life) began to popularise R_0 to represent the basic reproductive rate.

The story of R begins not with Covid-19 but with a very different disease: malaria

The Covid-19 R number: uses and limitations

Uses

Covid-19 and malaria are obviously very different diseases. Malaria is spread by a "vector" (mosquito bites) while Covid passes, usually airborne, directly from human to human. But the basic definition of R for any disease remains similar to Macdonald's. In its coronavirus style guide (tinyurl.com/2nz23cyt), the UK government defines it as "the average number of secondary infections produced by a single infected person".

As per Macdonald, 1 is the magic R number. The UK's chief medical officer, Chris Whitty, told the House of Commons Science and Technology select committee in April 2020 that:

> If R is 1, on average one person is giving this disease … to one person, and it is stable in the population. If R is 2, one person gives it to two people, who give it to four people, who give it to eight people. It is exponentially growing if it is anything above 1. If it is below 1 … it is falling away. Left to its own devices, if we did nothing, the R would naturally go above 1 again. (tinyurl.com/4u8nh9rz)

Rosalind Eggo, part of the modelling team at LSHTM, puts it even more simply: "Above 1, you have a problem – below 1, that's good". R tells you whether (and how quickly) your epidemic is growing, but is also useful "for giving a rough estimate of how much work you need to do to bring things down", according to Adam Kucharski – an R of 2 means you need to get transmission events down by half; an R of 1.5, by a third.

"R" may actually refer to at least three different numbers. R_0 (R-nought) is the natural reproduction number for the disease – the initial number at the start of an outbreak, when the population has no immunity. As the Covid-19 pandemic has gone on, R is much more likely to mean R_e or R_t, the effective reproduction number (or "Reff"). This is the R at a given moment in time, changing as governments enact measures to prevent the disease from spreading, the population changes its behaviour, and more of the susceptible population becomes infected, recovers, dies, or becomes immune through vaccination (and possibly becomes susceptible again). R_t can also refer to R excluding immunity – the estimate of growth given current transmission levels and taking into account social contact and restrictions, but assuming nobody has immunity. (Confusingly, the growth rate of a disease – the change in the number of infections from day to day – is sometimes written as *r*, or "little r".)

As GOV.UK, the UK government's official website, tells us, R "cannot be measured directly so there is always some uncertainty" (tinyurl.com/yc4dyhr6) – but it can be estimated using data. Key factors in calculating R are sometimes referred to as "DOTS" – the duration of time for which a person is infectious, the opportunities they have to spread it, the probability they will transmit it, and the susceptibility of the population. This can be worked out from observing the disease in patients, from case numbers, from deaths and even from data about how many people you might expect the average person to interact with.

R is useful because it is "one proxy answer" for "the missing statistic of the pandemic – how many people have Covid", according to Andrew Engeli, former deputy director at the UK Health Security Agency (which took on responsibility for producing R in July 2021). And many think it a useful and intuitive proxy – it gives a summary of the state of the epidemic and an expectation of the cases and deaths that will follow.

Understanding R

That does not mean R comes without confusion. Imagine a puddle in a football stadium doubles in size every day. After 10 days, the puddle has filled half the stadium. How many days will it take for the puddle to fill the whole stadium? Many people say 20 days – their version of doubling is geometric, to multiply the 10 days by 2. But the correct answer is 11 days, because the puddle is exponential – it doubles in size *every* day.

"The human brain prefers to think geometrically, not exponentially", says Fliss Bennee, co-chair of the Welsh Government's Technical Advisory Cell and Technical Advisory Group through much of the pandemic. "Without having a set of visual metaphors, it's almost impossible to explain how time kind of packs itself in into one end of an exponential [growth]". She puts a version of the puddle problem to policy-makers, with the puddle doubling in size every 5 seconds. The first few minutes, "it's a puddle, it's nothing". But soon, the stadium is full – and just 5 seconds before it's full, it will only be half full. "By the time [policy-makers] see what we can see, it will be too late to stop quite a lot of harm."

While a lot of the public have heard of R – in January 2021, MHP/Savanta ComRes found 88% of people had heard of the R number (tinyurl.com/y8kechuw) – not all of them understand it: 42% said they could confidently explain it whereas 46% could not; see Figure 1. Similarly, Ipsos MORI found that while two-thirds of the public said they were very or fairly confident they could understand Covid-19 statistics, only 47% could define the R number (tinyurl.com/bdct2cxn); see Figure 2. This level of knowledge and understanding is still impressive for an epidemiological statistic unknown by the general public a year before.

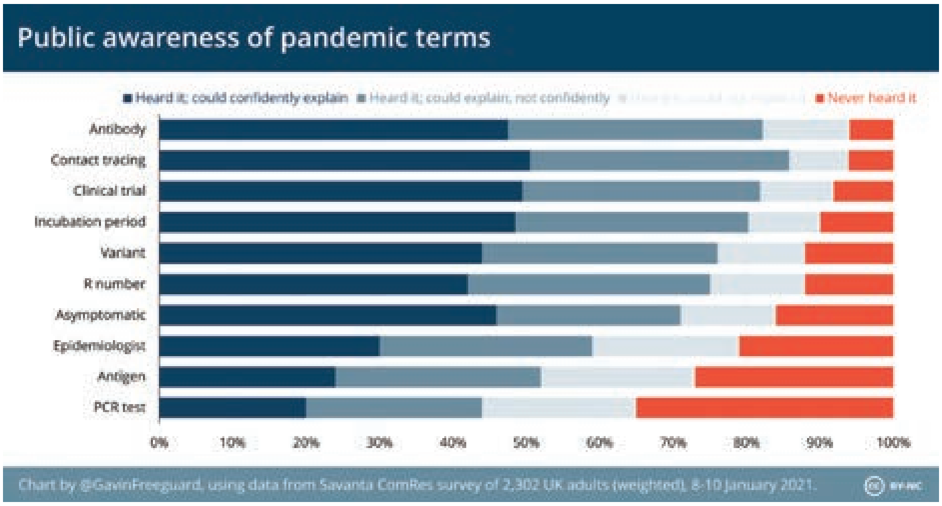

Figure 1: Public awareness of pandemic terms. Chart by Gavin Freeguard, using data from Savanta ComRes survey of 2,302 UK adults (weighted), 8–10 January 2021.

Limitations

R has its limitations. Because of the data it relies on, like cases, deaths and surveys, it is difficult to measure in real time and comes with a lag. As Rob Challen, a member of the modelling team at the University of Exeter, puts it: "By the time you take into account the Friday publication [of the R number by the UK government], from data produced on Monday, referring to the previous Thursday, referring to infections which happened … even with the best will in the world, you're generally looking at two weeks ago".

R is an average of many forecasts and usually at a high, often national level. That can mask variation between different parts of a country, and different communities and settings within it. Mark Woolhouse told the Lords Science and Technology Select Committee that R "is not the right thing to look at" because while R was coming down in the community, it was not doing so in care homes, "where the public health risks lie" (tinyurl.com/yczcvb6u). Early on in the UK pandemic, there was concern that particular ethnic groups were suffering – a letter to the *British Medical Journal* noted that "while those of black ethnicity comprise 3.3% of the population, they represent over 9.9%

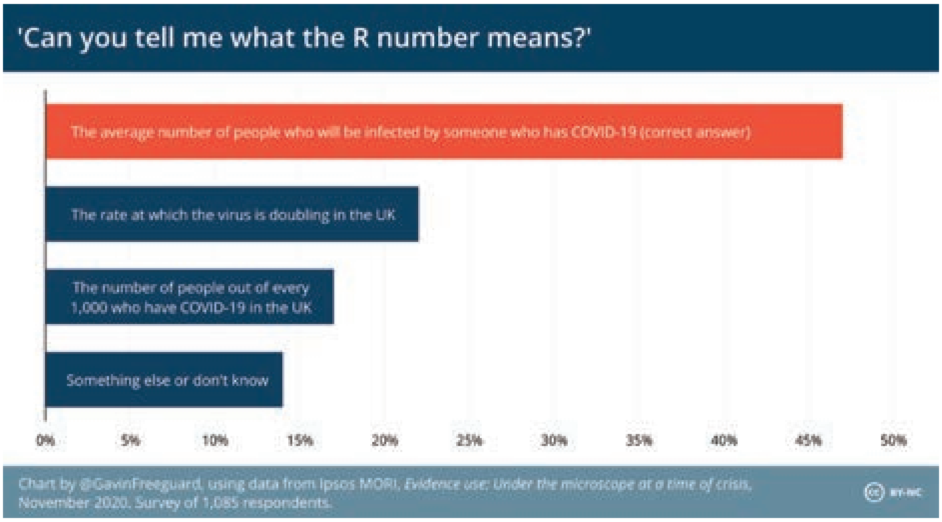

Figure 2: Answers to the question "Can you tell me what the R number means?". Chart by Gavin Freeguard, using data from Ipsos MORI survey of 1,085 respondents, November 2020.

of those critically ill" and that "calls into question … [R's] utility as a metric upon which important political decisions are made" (tinyurl.com/3vd6acrv).

In June 2020, the Commons Science and Technology Committee asked Sir Patrick Vallance, the government's chief scientific adviser, whether we might "get to a position where one person is infected with Covid in the country and that person infected five people – perhaps they are on the Isles of Scilly", giving a national R number of 5. He confirmed this was the case (tinyurl.com/5n88wpzx). This happened in Germany in June 2020: the national R rose from just over 1 to 2.88 because of a single large outbreak in a meat-processing plant in North Rhine-Westphalia. R can remain above 1 even while infections are decreasing.

The UK abandoned a UK-wide R in April 2021 in favour of regional estimates – but these also come with problems. As publications by the Scientific Pandemic Influenza Group on Modelling, Operational sub-group (SPI-M-O, a key subcommittee of the government's Scientific Advisory Group for Emergencies, SAGE) make clear: "When the number of cases falls to low levels and/or there is a high degree of variability in transmission across a region, then estimates of R and the growth rate become insufficiently robust to inform policy decisions" (tinyurl.com/3b2dw3m2).

A final problem is R being a victim of its own success, leading to an overreliance on it. Samir Bhatt, a professor of statistics and public health at Imperial College London, says: "Condensing a complicated situation to any one number is fundamentally flawed … A perfect one-number summary simply doesn't exist." Meaghan Kall, who led the Covid-19 epidemiology cell at Public Health England (now the UK Health Security Agency), says: "when you're creating a single indexed metric or a modelled metric like an R number that relies on many different data sources, that number inherits the limitations of all the different data sources". One "easily digestible" number that allows clear communication is "great" but "it takes a lot of the nuance out of the data that feeds it."

"R is one component of a complex system – no single component of which will give you the complete answer", Sir Jeremy Farrar, chief scientist of the World Health Organisation (and previously director of the Wellcome Trust, which funded this article) tells me. "It's about seeing the whole oil painting, not just one little bit of it."

GOV.UK – the UK government website – is open about R's shortcomings. The UK government publishes R as a range, not a single number: as Ed Humpherson, head of the Office for Statistics Regulation, told Parliament, this was "appropriate in conveying fuzziness, and the degree of uncertainty". The scientific publications and dedicated R number page on GOV.UK outline its limitations at length, and underline that both R (and the growth rate, published alongside it) "should be considered alongside other measures of the spread of disease, such as the number of people currently infected" (tinyurl.com/yc4dyhr6)

R will also be less useful at particular points in a pandemic. In response to their sensible Scilly questioning, Patrick Vallance told the select committee that R "was the right thing to measure" early on, but less so with cases falling. Fliss Bennee wondered how to explain to politicians and policy-makers that "there will be multiple waves when we will have to care about the R number more than anything else, until it gets to this point on the curve [when case numbers have fallen], when you need to focus on other things because it won't be valuable".

But for all its limitations, it is still useful. As Professor Dame Angela McLean, a member of SPI-M-O and now government chief scientific adviser, told one parliamentary select committee, having R means "the country does not spend hours and hours every week or every day saying, 'What size is the R number?'" – the consensus means we can "decide what we are going to do", rather than spending "so long arguing about it that terrible things happened while we were arguing" (tinyurl.com/4tsp8cde). ■

Next issue

In part two, we will look at how the UK's modelling groups tackled the Covid-19 R number.

Thanks to Understanding Patient Data (understandingpatientdata.org.uk) who first commissioned this text.

References

1. Macdonald, G. (1949) The distribution of epidemic diseases. *Health Education Journal*, **7**(2), 89–92.
2. Ross, R. (1910) *The Prevention of Malaria*. London: John Murray.
3. Kucharski, A. (2021) *The Rules of Contagion: Why Things Spread – and Why They Stop*. New York: Basic Books.
4. Macdonald, G. (1952) The analysis of equilibrium in malaria. *Tropical Diseases Bulletin*, **49**(9), 813–829.
5. Macdonald, G., Cuellar, C. B. and Foll, C. V. (1968) The dynamics of malaria. *Bulletin of the World Health Organization*, **38**, 743–755.