SIGNIFICANCE

February 2021 volume 18 issue 1

One year on

The toll

Excess deaths in the USA, Russia and elsewhere

The impact

A biostatistician's story of New York's deadly first wave

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

2021 CONFERENCE ON STATISTICAL PRACTICE

FEBRUARY 17–19, 2021

JOIN US ONLINE

Do you use statistics and data science in your daily work solving real-world problems? Want to communicate and collaborate more effectively with nonstatisticians and hone your skills?

The Conference on Statistical Practice brings together statistical practitioners—including data analysts, researchers, and scientists—with these goals in mind.

Short courses, tutorials, and sessions are designed to sharpen a broad spectrum of necessary skills in the following areas:

- **Leadership**
- **Communication**
- **Data Management**
- **Study Design**
- **Career Development**
- **Data Analysis**
- **Collaboration**
- **Big Data**

Learn more at ***ww2.amstat.org/csp***.

SIGNIFICANCE February 2021 volume 18 issue 1

NOTEBOOK

FEATURES

PROFILES

PERSPECTIVES

SIGNIFICANCE (1740-9705 and 1740-9713) is published bimonthly on behalf of the Royal Statistical Society by Wiley Periodicals LLC, 111 River St., Hoboken, NJ 07030-5774 USA. Periodicals Postage Paid at Hoboken, NJ and additional offices. Postmaster: Send all address changes to SIGNIFICANCE, Wiley Periodicals LLC, C/O The Sheridan Press, PO Box 465, Hanover, PA 17331 USA.

For ordering information, claims and any enquiry concerning your magazine subscription please go to wolsupport.wiley.com, email cs-journals@wiley.com or contact your nearest office.

Americas: +1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada). **Europe, Middle East and Africa:** +44 (0) 1865 778315. **Asia Pacific:** +65 6511 8000. **Japan:** For Japanese speaking support, email: cs-japan@wiley.com.

Statisticians call for "rigorous evaluation" of Covid test strategy in English schools

Mass testing may miss "many" coronavirus cases and give false reassurance, warn authors of *BMJ* article

nito/Bigstock.com

Statisticians have warned that a mass testing regime for Covid-19 in schools in England risks "spreading the disease more widely" and may lead to "even more disruption to education", following a year in which educational settings were closed to most pupils for several months.

The Department for Education (DfE) was previously advising that once in-school tests were rolled out, close contacts of a confirmed positive case in school would not need to self-isolate, so long as those close contacts remained symptomless, agreed to daily testing for a period of seven days, and returned negative test results (bit.ly/38XgEee). The government's hope was that this would avoid uninfected people having to self-isolate, meaning it could keep as many staff and pupils in school and college as possible.

However, in an opinion piece published in the *British Medical Journal* (*BMJ*), a group of statisticians expressed concern that the DfE's plans may have unintended consequences (bit.ly/39RHFiO). They argued that negative results from the tests that schools are being told to use – lateral flow tests – are "too inaccurate to rule out" Covid, meaning that "The possibility that some close contacts who are infected will test negative and will spread the virus is not negligible."

On 20 January, the DfE changed tack and announced that daily contact testing in schools would be "paused", and that contacts of a confirmed case would have to self-isolate in line with general public health guidance (bit.ly/2LH2Dc8). In the meantime, the DfE said, experts will look at "whether daily testing is effective" given that a new, more transmissible strain of the virus has become "dominant".

The DfE stated that the second arm of its in-school testing strategy – twice-weekly testing of staff to identify asymptomatic cases – would continue. However, the authors of the *BMJ* article had also expressed reservations about this part of the strategy, warning that tests may miss "many" such cases "and falsely reassure those testing negative".

The article, by Jon Deeks, Mike Gill, Sheila Bird (a member of the *Significance* Editorial Board), Sylvia Richardson and Deborah Ashby, cites several studies that compare the results of lateral flow tests to those of PCR tests done at the same time (PCR is described by the authors as the "gold standard"). Comparisons were made of tests conducted in various settings – from tests of symptomatic patients in hospitals, to mass testing of symptomless people in the city of Liverpool and among University of Birmingham students.

The *BMJ* authors – several of whom are members of the Royal Statistical Society's Covid-19 Task Force or its Diagnostic Test Working Group – report that studies of symptomatic patients "show test performance declining when not done by experts, as will happen in schools", and that test performance was "worse" in studies of symptomless people.

However, in a Twitter thread on 21 January, Public Health England pointed to "new research which finds that these tests are most effective in detecting people with high viral loads, who are most likely to pass the virus on to others" (bit.ly/39Tg6p7).

Educational settings were closed to most pupils and students in January as part of a new national lockdown – England's third – which is expected to last until at least the February half-term holiday. Some schools have started using lateral flow tests already. But Deeks and colleagues say that any wider implementation of the mass testing programme should not happen without "rigorous evaluations" to compare the current proposed strategy with other testing options. ■

Met Police to collect ethnicity data on vehicle stops

Six-month pilot follows "concerns about racial profiling and disproportionality"

London's Metropolitan Police Service (MPS) has started a six-month pilot project to collect data on the ethnicity of drivers stopped by officers. The pilot was called for in a November 2020 action plan published by the Mayor of London, Sadiq Khan, and aims to "identify any disproportionality relating to ethnicity".

According to the mayor's action plan, ethnicity data is already captured when a search of a person or vehicle is made following a vehicle stop (bit.ly/3nWfX9s). "This data shows that Black people are six times more likely than white people to be stopped and searched", reads the action plan. However, currently there is no requirement to record ethnicity for vehicle stops that do not result in a search. This is described as "a blind-spot that must be resolved".

"[T]his pilot will help us to begin to assess and address concerns about racial profiling and disproportionality in our city," said Khan in a statement (bit.ly/2XSIvGs). "Road Traffic Stops are an important tool the police have to keep Londoners safe but they can have a huge impact on community relations and deserve the same level of scrutiny as any other kind of police stop-and-search power."

The mayor says that he has written to Home Secretary Priti Patel "to ask her to make it compulsory for the police to collect and publish data on ethnicity for Road Traffic Stops because it is absolutely vital that our police service retains the trust and confidence of all the communities it serves."

Khan's action plan was drawn up in the wake of the Black Lives Matter protests of 2020. According to the mayor's office, the plan was informed by consultations with "more than 400 individuals and groups that either work with or within Black communities" (bit.ly/30W7rIJ).

In announcing the pilot, the MPS explained that officers across London "will record the location and time of the vehicle stop, ethnic background, sex and age of the driver, and the make and model of the vehicle" (bit.ly/3io6pCS). Data will be recorded at the end of a vehicle stop, after policing duties have been carried out. ■

Write here, write now

Enter the 2021 Statistical Excellence Award for Early-Career Writing

Data took centre stage in 2020 as the world was gripped by the Covid-19 pandemic. In 2021, and with the pandemic still raging, we want to shine a spotlight on data stories from around the globe. So, we are pleased to once again be organising the Statistical Excellence Award for Early-Career Writing, in partnership with the Young Statisticians Section of the Royal Statistical Society.

This international award celebrates career-young statisticians, data scientists and researchers who can demonstrate the skills necessary for effective communication and who recognise the importance of explaining statistics to non-experts.

The rules of entry are simple: competition entrants are invited to submit their best statistical writing in the form of a magazine article (1,500 to 2,500 words) on any subject they like. Articles will be reviewed by a judging panel, and the winning entry (and up to two runners-up) will be published by *Significance* later this year.

Successful submissions from previous years have been based on original analyses produced specifically for the competition. Past participants have also written about work they have done as part of their studies or during their careers, while some have written about the work of others in the form of a critique or wider overview of a subject area.

Gift Habeshaw/Unsplash.com

Whatever the topic, articles must be engaging and easy to read. *Significance* is published for a broad audience, so accessibility is key. And articles must be submitted by the deadline, 31 May 2021. See significancemagazine.com/writingcomp for full details. ■

US protests and police use of force

An analysis of data on law enforcement responses to protests across the United States since April 2020 suggests that police "are three times more likely to use force against leftwing protesters than rightwing protesters", according to *The Guardian* (bit.ly/38TAgjp). Citing statistics from the US Crisis Monitor database, *The Guardian* reports that 4.7% of leftwing protests resulted in the use of force by law enforcement, compared to 1.4% of rightwing demonstrations. Comparing only peaceful protests, 1.8% of leftwing demonstrations were met with force versus 0.5% of rightwing protests. US Crisis Monitor is online at bit.ly/3oVhgXq.

US Census Bureau director steps down

Steve Dillingham, the director of the United States Census Bureau, retired from his position on 20 January 2021. He had almost a year left to run before his term was due to end. In a blog post, Dillingham said his "planned departure would have occurred earlier, but I received requests to continue serving during and after the transition" to President Biden's administration (bit.ly/3oWho9j). "But I must do now what I think is best," said Dillingham. Referring to the 2020 Census, he added: "I fully expect that President-Elect Biden will have complete confidence in the results that he will announce."

US election polls: a quick postmortem

How did the 2020 US presidential election polls really do? **Ole J. Forsberg** gives his assessment

The American Association for Public Opinion Research (AAPOR) is expected to produce a report early this year that explores the strengths and weaknesses of the polls in the 2020 US election cycle. The polls were criticised in some quarters immediately after the election, when it became clear that Donald Trump had done better than expected and that Joseph R. Biden Jr's margin of victory in the popular vote was not as large as anticipated.[1]

In preparation for this report, I wanted to provide some insight into the polls and some suggestions of my own for moving forward. Specifically, I hope to convince polling houses to use some type of model averaging – or even Bayesian methods – to reflect uncertainty in the voting population, and to encourage better explanations of poll results to the media and their readers.

Comparing 2016 and 2020

I expect that the AAPOR report, when published, will likely focus on the same three sources of error that were discussed in its May 2017 report covering the 2016 US election polls (bit.ly/3ihEYuH). According to that report, the 2016 polls underestimated Trump's eventual support because of (1) a failure to properly weight for the education level of respondents, (2) "shy Trump voters" outnumbering "shy Hillary Clinton voters" (either in response or non-response), and (3) a genuine shift in voter preference during the closing weeks of the campaign.

The first source of error, faulty weighting, is extremely important for polling houses to take seriously. While the number of US polling houses taking education level into consideration increased in 2020, the education characteristics of the voting population remain uncertain.

"Shy voters" – the second source of error – may be more myth than reality (53eig.ht/30NEb6R). But whether shy or not, there are some voters who either choose not to respond to polls, or who choose not to answer honestly when surveyed. Pollsters need to address this, either by asking additional questions to model respondent preference for those who choose not to say how they will vote, or by finding new ways to encourage the public to respond to legitimate polls, or even by using the non-response rate as an indicator of greater uncertainty in polling estimates.

I contend that the third source of error – a late shift in voter preference – is an error of interpretation, not of polling. The mistake happens in how we interpret a poll result such as "48% Biden, 44% Trump". Do we focus on the two-party vote and claim that Biden is ahead, or do we acknowledge that there is a sizeable portion of voters – 8% – who may only decide how to vote once in the polling booth? Clearly, the latter interpretation is more appropriate, but it makes for a less straightforward story, so these undecided voters tend to be overlooked in media reports.

Missing data

The majority of polls in the 2020 election cycle contained just three response options for those asked about their intended vote: "Biden", "Trump", and "undecided". The implied fourth option was "I refuse to take this poll" – and about 90% of people chose that "option" when contacted by a polling house (response rates were below 10%).

Taken together, these non-respondents and the undecideds mentioned earlier constitute a huge amount of missing data about voting intention. Ignoring these missing data leads to false precision in the polls' assessment of the state of the election.

While some undecided voters ultimately will not vote, many will eventually decide between the two candidates. This increases the uncertainty in polling estimates beyond what is reported in terms of confidence intervals and margins of error. As a result, when those late-deciding voters finally vote, polls may look very wrong.

To illustrate this point, compare the polls in the final two weeks of the 2020 election to the final election result (Table 1). In this sample of 174 polls, the actual Biden vote was within the polls' margins of error 85% of the time, while the actual Trump vote was within the polls' margins of error only 43% of the time. For the 57% of confidence intervals that missed Trump's actual vote, they were always too low, never too high – meaning that the polls consistently underestimated Trump's final vote. The 15% of confidence intervals that missed Biden's actual vote were roughly balanced between those that were too high and those that were too low. In other words, the polls tended to do a much better job of estimating the Biden vote than the Trump vote. But,

Table 1: Results from comparing candidate support levels in polls from the last two weeks of the US presidential election with the actual outcome of the election (vote share). Polls are a mix of state-level and national polls from a variety of polling houses, using a variety of methods.

		Confidence interval hits		Average miss (standard error)	
Source	***n***	**Biden**	**Trump**	**Biden**	**Trump**
All polls	174	85% (79% to 90%)	43% (35% to 50%)	–0.09	+2.41
Online only	23	96% (78% to 99%)	30% (13% to 53%)	–0.79	+2.21
Online + telephone	26	92% (75% to 99%)	54% (33% to 73%)	–0.78	+2.24
Telephone only	125	82% (74% to 88%)	42% (34% to 52%)	–0.18	+2.48
University	60	92% (82% to 97%)	27% (16% to 40%)	–0.10	+2.99
Non-university	114	82% (73% to 88%)	51% (41% to 60%)	–0.09	+2.10
Partisan	52	79% (65% to 89%)	75% (61% to 86%)	+0.62	+1.33
Non-partisan	122	88% (81% to 93%)	29% (21% to 38%)	–0.40	+2.87

Ole J. Forsberg is an assistant professor of mathematics-statistics and the director of the statistics program at Knox College in Illinois. He specialises in modelling and testing elections and is the author of *Understanding Elections through Statistics: Polling, Prediction, and Testing*.[2]

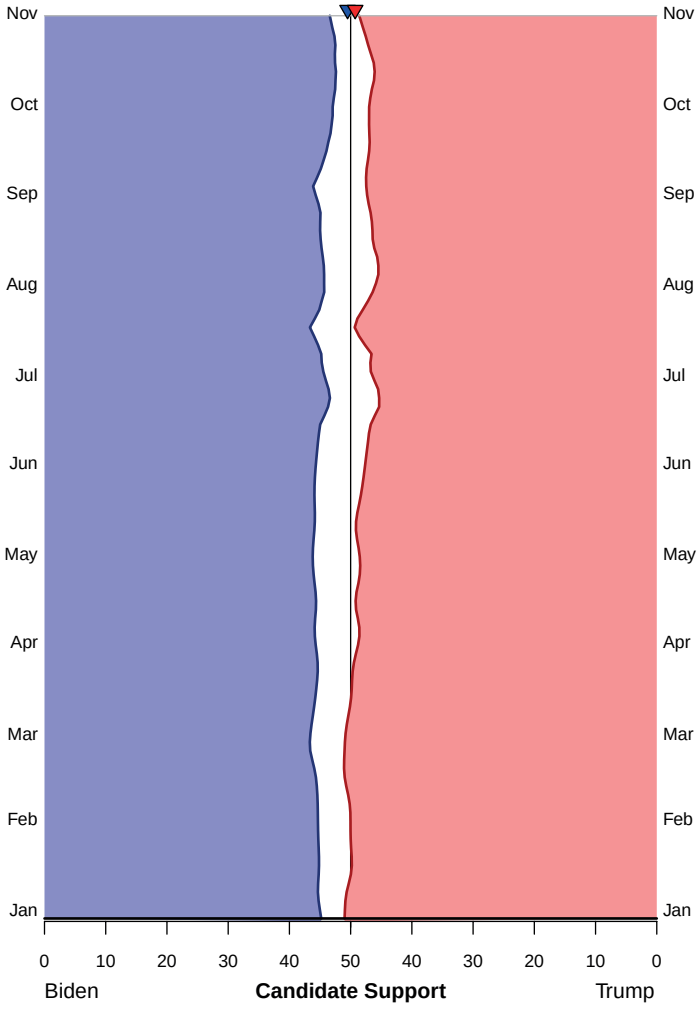


Figure 1: Estimated US presidential vote over time for Georgia, January to November 2020. The blue curve is the estimated support for Biden; the red curve, for Trump. The space between the two represents the proportion of undecided voters. The election results are indicated with the triangles at the top: Biden at 49.5%, Trump at 49.3%.

even so, they offered no clue as to how undecided voters would eventually vote.

Figure 1 illustrates this problem again, this time specifically for the state of Georgia. The left curve is the estimated support for Biden over time; the right, for Trump. The gap between the two curves represents the estimated proportion of undecided voters on any given day. That the election outcomes (triangles at the top) sit within this gap suggests that the polls did quite well in estimating each candidate's core support. But they failed in estimating how the undecideds would break on election day.

The unknown population

Some will of course argue that it is not the job of polls to predict how people will vote, especially those who are undecided. Polls exist simply to offer a snapshot of how people say they intend to vote at a given point in time, based on a representative sample of the voting population. But here is where it gets tricky: the voting population does not exist until election day. There is a population of *eligible* voters before election day, but not all eligible voters vote. This means that polling houses must estimate the characteristics of the voting population in order to recruit and weight their samples. They may use characteristics such as gender, political orientation, wealth, age and – yes – educational attainment. But because politics is not static, the characteristics used should be dependent on the election and its features. Crucially, pollsters will not know whether their samples are based on the right mix of population characteristics until after the election is won.

Currently, polling houses tend to settle on a single weighting scheme (weights assigned to each stratum in a stratified sample) and apply it to their raw data to achieve their final estimates. However, it would be more statistically sound for polling houses to acknowledge the uncertainty in the expected voting population and incorporate this into their estimates. This could be as simple as using several different "voting populations" to create several estimates of "voter support", for which pollsters then report the average. It could be as sophisticated as using Bayesian methods to place a prior distribution on the population strata and reporting the posterior mean and credible interval.

Personally, I favour the Bayesian solution because it provides a solid statistical structure for estimation and communication of results. Using Bayesian methods would force pollsters to acknowledge yet another source of uncertainty in their estimates and this may, in turn, encourage pollsters to be more modest with their results when communicating with the media. Such an approach may also help the media to better understand the inherent uncertainty in poll results so that they can convey this to their readers, viewers, and listeners.

Communication

This leads me to what I think is a key lesson to be learned from the 2020 election polls. The end-user, the typical media consumer, tends not to have a solid understanding of statistics. Furthermore, they may not have the time to learn about statistics and what the poll numbers really mean. This places an additional burden on pollsters to ensure their results – their estimates, their uncertainties, *and* their meanings – are properly reported.

My view is that the polls, overall, did quite well. However, some media reports throughout the campaign failed to communicate what the polls were actually saying. Those same reports also failed to explain what polls are even *capable* of saying.

Polls provide a tantalising glimpse into the current state of some unknown future population. The presence of undecided voters adds to this uncertainty. If pollsters were to better convey this uncertainty and all that it means, it may lead the media to report polls differently, which may help to create reasonable expectations in future of what polls can and cannot tell us. ■

Disclosure statement

The author declares no competing interests, financial or otherwise, relevant to the content of this submission beyond the author's academic appointments.

References

1. Tarran, B. (2020) US election called for Biden, as polls face criticism. *Significance*, **17**(6), 2.

2. Forsberg, O. J. (2021) *Understanding Elections through Statistics: Polling, Prediction, and Testing*. Boca Raton, FL: CRC Press.

How to lose weight well, according to *How to Lose Weight Well*

Authorities in the UK hope to "tackle obesity" by encouraging people to "eat better and move more". But what sort of dietary changes are most effective? **Joshua E. Stubbs** and **Toby C. T. Stubbs** look to a TV show for answers

According to Public Health England, being obese notably increases the risk of experiencing hospitalisation and death from Covid-19 (bit.ly/3aqRbLt). In light of this, the UK government has recently announced a strategy to reduce obesity (bit.ly/3ancNIA). Globally, the age-standardised prevalence of obesity among adults has increased 1.5 times over the past two decades.[1] Describing Covid-19 as "a wake-up call", the UK government says: "We need to use this moment to kick start our health, get active and eat better."

But what sort of dietary changes are most effective at helping to shift excess weight? For the past five years, a British TV channel, Channel 4, has been airing a programme called *How to Lose Weight Well*, which aims to "road-test" a variety of diets to help viewers make better-informed decisions about which are worth considering.

In each episode, three pairs of family members or friends adopt diets for a set period. The first pair, *crashers*, adopt diets of approximately 1–2 weeks; the second pair, *shape shifters*, adopt diets of 4–6 weeks; and the third pair, *life changers*, adopt diets of 12–16 weeks. Crashers typically adopt the most extreme diets, which can include, for example, consuming baby food or bone broth as meal replacements. Viewers are advised that rapid weight loss can be dangerous, difficult to maintain, and that medical advice should be sought before attempting to lose weight quickly. Shape shifters, in contrast, typically adopt less extreme but nonetheless challenging diets, such as Dr Michael Mosley's blood sugar diet,[2] which involves consuming no more than 800 calories each day, as well as fewer carbohydrates and more Mediterranean-style food. Life changers typically adopt the least extreme diets, such as time-restricted eating, which involves only consuming calories within a set period (between 10 a.m. and 8 p.m., for example).

It is often unclear which diets are most effective, however. In the most recent episode, for example, shape shifters and life changers lost a similar amount of weight. Watching a handful of additional episodes does not help much, either, because it yields a small and heterogeneous sample of dieters and diets from which it is difficult to generalise. We therefore attempted to identify which diets were most effective by comparing the dieters' baseline and post-diet weight after watching all five seasons of *How to Lose Weight Well* and recording details about the different diets that the 137 dieters adopted (one dieter had to abandon their diet for health-related reasons).

How to Lose Weight Well is a TV programme that aims to "road-test" a variety of diets to help viewers make better-informed decisions about which are worth considering

Weights and weight lost

Measured in pounds before beginning their diets, crashers (mean = 188.7; standard deviation (SD) = 31.3) weighed substantially less, on average, than both shape shifters (mean = 221.9; SD = 44.8) and life changers (mean = 227.5; SD = 43.0). Almost two-thirds (87) of the dieters were female; two-fifths (53) had to stop consuming alcohol; and approximately a quarter had to calorie-count (40) or were obliged to exercise (30). Less than a fifth had to stop consuming caffeine (22) or adopt a vegan diet (13).

The percentage of body weight lost by diet group is displayed in Table 1, Table 2 and Figure 1. Almost all dieters (136) lost weight, with two-thirds (90) losing at least 5% of their body weight and a quarter (32) losing at least 10%.

Differences between the diet groups are apparent, however. On average, life changers lost a much higher percentage of their body weight than shape shifters, who in turn lost more than crashers (see "Statistical tests and results").

While less than two-fifths of crashers (17) lost more than 5% of their body weight, in excess of three-quarters of both shape

Table 1: Measures of central tendency and dispersion in the percentage of body weight lost by diet group.

	Percentage of body weight lost			
	Mean	Median	SD	Range
Crashers (46)	4.6	4.4	2.3	10.4
Shape shifters (46)	6.6	6.5	2.8	14.3
Life changers (45)	10.9	10.8	5.6	19.4
Total	7.3	5.8	4.6	22.3

Table 2: Percentage of body weight lost by diet group.

	Percentage of body weight lost			
	Crashers (46)	Shape shifters (46)	Life changers (45)	Total
0–4.9%	63.1	21.7	15.5	33.5
5–9.9%	34.7	63.0	28.8	42.3
10–14.9%	2.2	13.0	31.1	15.3
15–19.9%	0.0	0.0	17.7	5.8
20–24.9%	0.0	0.0	6.6	2.1

Note: one *shape shifter* gained weight.

Joshua E. Stubbs is a PhD education student in the Psychology in Education Research Centre (PERC), Department of Education, University of York.

Toby C. T. Stubbs is a BSc psychology student in the School of Psychology, University of Leeds.

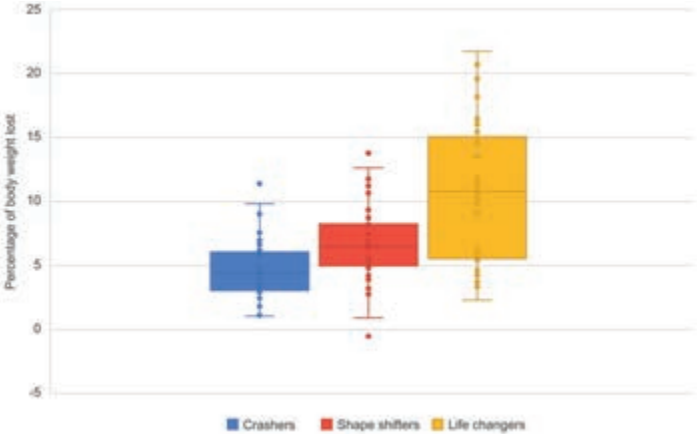


Figure 1: Percentage of body weight lost by diet group.

shifters (35) and life changers (38) did so. Differences between shape shifters and life changers are more perceptible when the proportion of those who lost at least 10% of their body weight is examined; while less than a sixth of shape shifters (6) lost more than 10% of their body weight, more than half of life changers (25) did. Furthermore, while none of the crashers or shape shifters lost more than 15% of their body weight, one quarter of life changers (11) exceeded this cut-off.

It is therefore clear that the longer the diet, the more likely the dieter was to lose a large(r) proportion of their body weight, and that the length of diet was typically more important than the extremity. Indeed, a Pearson correlation indicated that there was a moderate positive correlation between diet length and percentage of body weight lost ($r(136) = 0.57$, $p < 0.001$), even though the most intense diets were also the shortest. When those on the shortest diets of approximately 1–2 weeks (i.e. crashers) were removed from the analysis, a Pearson correlation still indicated that there was a moderate positive correlation between diet length and percentage of body weight lost ($r(90) = 0.45$, $p < 0.001$).

Different diets

Percentage of body weight lost according to what type of diet the dieters adopted was also analysed (and these data can be viewed online as Table 3 at significancemagazine.com/694). Due to the small sample sizes that result from disaggregating dieters according to the type of diet they adopted within their respective diet groups, a high degree of caution should be taken when attempting to make inferences about which types of diets are most effective. It is particularly important to exert a high degree of caution when considering the role that making exercise obligatory performed, because many dieters who were not obliged to exercise clearly did so with greater frequency and intensity when dieting.

With these cautions in mind, no particular type of diet within this data set stands out as the most effective. The most perceptible differences in the percentage of body weight lost are to be found when comparing the diet length rather than whether it involved calorie counting, abstaining from alcohol or caffeine, or adopting a vegan diet.

Statistical tests and results

Given the high degree of similarity between the baseline weight of shape shifters and life changers, an independent-samples *t*-test was conducted to compare the average percentage of body weight lost between these groups. The results of the test indicate that the two groups differ in the percentage of body weight lost, and that there is a large effect size ($t(89) = 4.6$, $p < 0.001$, $d = 0.97$).

Independent-samples *t*-tests assume that the variances in the scores for the variable being compared are approximately equal. So, a Levene test was conducted to evaluate the variances in the percentage of body weight lost by shape shifters and life changers. As this test indicated unequal variances in the percentage of body weight lost by these two groups ($F = 22.3$, $p < 0.001$), the degrees of freedom were adjusted from 89 to 65.

The longer the diet, the more likely the dieter was to lose a large(r) proportion of their body weight

What have we learned?

Analysis of data from *How to Lose Weight Well* falls short of identifying which specific diets are most effective. Rather, the findings suggest that if we would like to lose a substantial amount of excess body weight, our best bet is to opt for sustainable diets that we think that we will stand a reasonable chance of maintaining for several months rather than weeks.

For viewers to be able to put the changes observed in each episode in greater context, future iterations of *How to Lose Weight Well* would benefit from communicating more clearly how much weight dieters typically lose, and for how long such weight loss is generally maintained. Indeed, a study recently published in the *British Medical Journal* found that while most dieters lose weight, regardless of what type of diet they adopt, dieters also tend to regain weight once their diet has finished.[3] It is therefore important that *How to Lose Weight Well*, alongside other programmes and campaigns that contribute towards the architecture of public health messaging, emphasise even more clearly the need for long-lasting lifestyle changes rather than short-term and faddish diets. ■

Disclosure statement

The authors declare no conflicts of interest or affiliations beyond their academic positions.

References

1. World Health Organization (2020) *World Health Statistics 2020: Monitoring Health for the SDGs, Sustainable Development Goals*. Geneva: WHO.

2. Mosley, M. (2015) *The 8-Week Blood Sugar Diet: Lose Weight Fast and Reprogramme Your Body*. London: Short Books.

3. Ge, L., Sadeghirad, B., Ball, G. D. C., da Costa, B. R., Hitchcock, C. L., Svendrovski, A. *et al.* (2020) Comparison of dietary macronutrient patterns of 14 popular named dietary programmes for weight and cardiovascular risk factor reduction in adults: Systematic review and network meta-analysis of randomised trials. *British Medical Journal*, **369**, m696.

Passionate statisticians

The phrase "passionate statistician" is usually reserved for Florence Nightingale. But she was not the first to be described as such, as **John Aldrich** explains

The phrase "passionate statistician" has come to personify Florence Nightingale (1820–1910). However, she did not refer to herself in such terms. Though she wrote of her "passionate study" in an 1872 letter to her statistical master, Adolphe Quetelet,[1] it was her biographer, Edward Tyas Cook (1857–1919), who applied the "passionate statistician" tag.

The main theme of Cook's two-volume biography, *The Life of Florence Nightingale*, is Nightingale's passionate nature.[2] His "passionate statistician" echoes through the *Life* and accumulates depth as the book progresses. Nightingale was not the first "passionate statistician", however. The phrase was first applied by the politician and banker George Joachim Goschen (1831–1907) to himself, before being picked up and used to comic effect by the satirical weekly *Punch*.

Goschen and *Punch*

The phrase in which Cook invested so much came ready-made from a speech by Goschen, who was Chancellor of the Exchequer between 1887 and 1892. The speech was reported in *The Times* on 9 March 1885 and was lampooned in *Punch* on 21 March. In the 1880s Goschen got a lot of attention from *Punch*, as had Nightingale in the 1850s (see Figure 1).

Goschen was addressing the annual meeting of the London Society for the Extension of University Teaching when he used the "passionate statistician" phrase. He was president of the society's council and the place was Toynbee Hall, a university settlement in Whitechapel, where students from Oxford and Cambridge lived, studied and did social work in the deprived areas of London. The extension and settlement movements were attempts to take university to the working class. Goschen – and Cook, who was in the audience – supported both.[3–5]

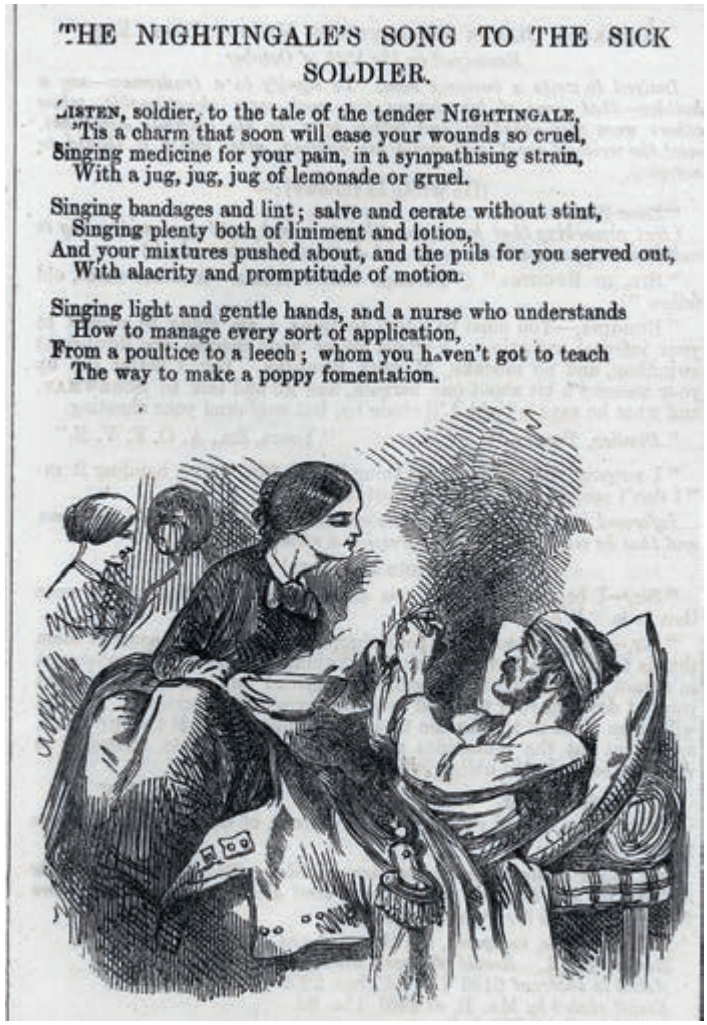
Goschen (and Cook) had studied classics at Oxford, but the extension classes were not restricted to traditional subjects. They aimed to develop intellect, and Goschen told his audience that there was "something in every subject that would give deep interest to the student", instancing the seemingly "dry" studies of statistics and butterflies. There was nothing funny about lepidoptery – butterfly-collecting was a well-regarded pastime – but there was amusement that "some people enjoyed the study of statistics". Goschen, as reported in *The Times*, continued:

> For his part, he was a passionate statistician. (Laughter.) ... There were a hundred interesting facts in economics, in national history, in the history of the world, which statistics would teach. If they would go with him into the history of statistics he would make them all enthusiasts in statistics. (Laughter.)

Goschen spoke as an educationalist and statistical hobbyist, but *Punch* revoiced his offer. Goschen's invitation became a poem, "The Passionate Statistician to His Love", modelled on Christopher Marlowe's "The Passionate Shepherd to His Love", a sixteenth-century work familiar to Victorian readers. As was customary in *Punch*, the poem was unsigned, but Richard Noakes has identified the author as Edwin James Milliken (1839–1897), who was for decades one of the *Punch* team.

Marlowe's "Passionate Shepherd" invites his beloved to share the delights of his life; the first verse (of six) is, "Come live with me, and be my love,/ And we will all the pleasures prove/That Valleys, groves, hills, and fields,/Woods, or steepy mountain yields." Milliken's "Passionate Statistician" instead begins: "Come live with me and be my love,/And we will all the pleasures prove/That facts and figures can supply/ Unto the Statist's ravished eye."

The poem (Figure 2) received favourable notice. But its fame was temporary, and the label – mocking or otherwise – did not become fixed to Goschen as it later did to Nightingale.



THE NIGHTINGALE'S SONG TO THE SICK SOLDIER.

LISTEN, soldier, to the tale of the tender NIGHTINGALE,
'Tis a charm that soon will ease your wounds so cruel,
Singing medicine for your pain, in a sympathising strain,
With a jug, jug, jug of lemonade or gruel.

Singing bandages and lint; salve and cerate without stint,
Singing plenty both of liniment and lotion,
And your mixtures pushed about, and the pills for you served out,
With alacrity and promptitude of motion.

Singing light and gentle hands, and a nurse who understands
How to manage every sort of application,
From a poultice to a leech; whom you haven't got to teach
The way to make a poppy fomentation.

Figure 1: Florence Nightingale reverentially portrayed by *Punch* in 1854.

John Aldrich is an emeritus fellow of the University of Southampton, working on the history of statistics and economics. He is chair of the Royal Statistical Society's History of Statistics Section.

Cook and Nightingale

Nightingale, Goschen and *Punch* offer three different takes on the passionate statistician, with Cook holding them together. Cook took an existing phrase, with which he had a personal association, and found a nice use for it.

In describing Nightingale as a "passionate statistician" in volume 1, chapter 2 of his biography of her, he also relates the story of Goschen and *Punch*. So far, so natural. Yet the Goschen–*Punch* story jars with the rest of the *Life*: Cook respected statisticians while *Punch* made fun of them, and Goschen fell far short of Nightingale, who was so much more of a passionate statistician.

Goschen was a politician who looked to numbers to support the work of government. Like several politicians, Goschen was in the Statistical Society of London (later the Royal Statistical Society), joining in 1868 and becoming president in 1886. Most were just well-wishers, but Goschen was more, and he had a pet statistical thesis – on the growth of middle incomes – that he expounded in his presidential address.[6] Though he was more than a hobbyist, Goschen did not have Nightingale's deep intellectual interest in statistics or the same passion for humankind. She met him in 1869 when he was president of the Poor Law Board, and Cook quotes from her assessment: "of considerable mind, great power of getting up statistical information and political economy, but with no practical insight or strength of character". Cook, who knew Goschen, thought this "a little severe, perhaps, but not undiscriminating".[2]

Cook's biography of Nightingale was a book of revelations, and Nightingale's statistical side was one of them. Her statistical thoughts and transactions with Quetelet and Francis Galton had never been public, and there was no memory of her public life as a statistician: this had ended decades earlier; her associates, William Farr and Thomas Graham Balfour, were long dead, and she had no young disciples. Cook reviewed Nightingale's statistical reports and noted her forgotten diagrams. He summed up:

> With Miss Nightingale statistics were a passion and not merely a hobby. They did, indeed, please her, as congenial to the nature of her mind. ... [S]he loved statistics, not for their own sake, but for their practical uses.

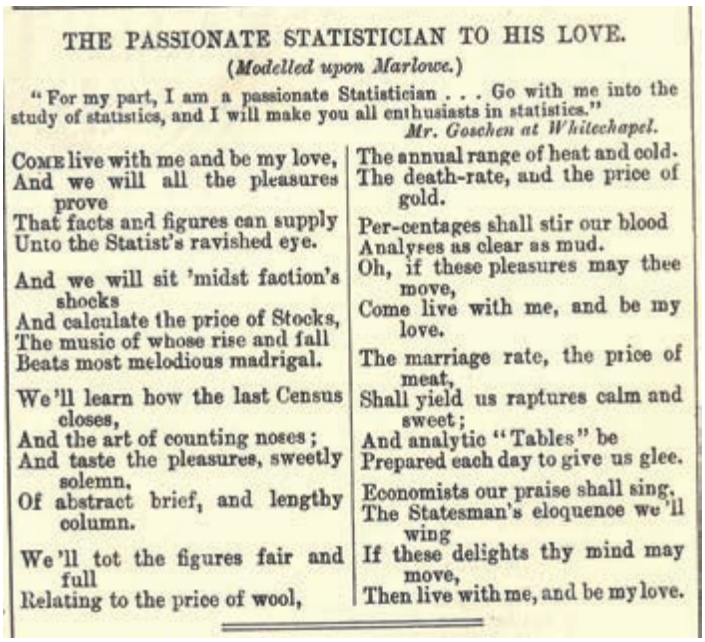
Chapter 5 revealed another side to Nightingale: the religious thinker and author of *Suggestions for Thought*, which set out a statistical worldview.[7]

In volume 2 of the *Life*, Cook drew on these discussions to explain how the mundane matter of Nightingale's enthusiasm for introducing statistics at Oxford – something she discussed with Galton – was connected to one of the "ruling thoughts" of the life of "a Passionate Statistician", duly capitalised. Cook explains that for her:

> The true function of theology was to ascertain "the character of God." Law was "the thought of God." It was by the aid of statistics that law in the social sphere might be ascertained and codified, and certain aspects of "the character of God" thereby revealed. The study of statistics was thus a religious service.

Cook's Nightingale had a range and depth of passion to make her unique but, though he used "The Passionate Statistician" as a chapter title, he did not assert that there was only one "passionate statistician": he acknowledged the existence of others – her allies in one parliamentary battle were "passionate statisticians", and sceptics about one of her schemes were "not passionate statisticians".

The label of "the passionate statistician" endures, fixed to Nightingale and applied to others only if they have Nightingale qualities – such as the American economist and social worker Edith Abbott (1876–1957).[8] The *Punch* poem appears to have been forgotten, although Irwin Collier found the text – retitled "Ode to an Economist" – when excavating a University of Chicago tradition of skits about economists (bit.ly/3i7kNiM). Skitting goes on and I had hoped to find new examples, but modern verse and pop songs do not seem to offer models of the passionate involver. ■



THE PASSIONATE STATISTICIAN TO HIS LOVE.

(*Modelled upon Marlowe.*)

"For my part, I am a passionate Statistician . . . Go with me into the study of statistics, and I will make you all enthusiasts in statistics."
Mr. Goschen at Whitechapel.

Come live with me and be my love,
And we will all the pleasures prove
That facts and figures can supply
Unto the Statist's ravished eye.

And we will sit 'midst faction's shocks
And calculate the price of Stocks,
The music of whose rise and fall
Beats most melodious madrigal.

We'll learn how the last Census closes,
And the art of counting noses;
And taste the pleasures, sweetly solemn,
Of abstract brief, and lengthy column.

We'll tot the figures fair and full
Relating to the price of wool,
The annual range of heat and cold.
The death-rate, and the price of gold.

Per-centages shall stir our blood
Analyses as clear as mud.
Oh, if these pleasures may thee move,
Come live with me, and be my love.

The marriage rate, the price of meat,
Shall yield us raptures calm and sweet;
And analytic "Tables" be
Prepared each day to give us glee.

Economists our praise shall sing,
The Statesman's eloquence we'll wing
If these delights thy mind may move,
Then live with me, and be my love.

Figure 2: *Punch*'s March 1885 poem, making fun of Goschen's claim.

Disclosure statement

The author declares no conflicts of interest.

References

1. Diamond, M. and Stone, M. (1981) Nightingale on Quetelet. *Journal of the Royal Statistical Society*, *Series A*, **144**(1), 66–79.

2. Cook, E. T. (1913) *The Life of Florence Nightingale*, 2 volumes. London: Macmillan.

3. Barnett, H. O. W. (1918) *Canon Barnett. His Life, Work, and Friends*. London: John Murray.

4. Saxon Mills, J. (1921) *Sir Edward Cook KBE: A Biography*. London: Constable.

5. Evans, R. A. (1982) The University and the City: The educational work of Toynbee Hall, 1884–1914. *History of Education*, **11**(2), 113–125.

6. Goschen, G. J. (1887) "The increase of moderate incomes;" being the inaugural address of the President of the Royal Statistical Society, the Right Hon. G. J. Goschen, M.P., Chancellor of the Exchequer. *Journal of the Royal Statistical Society*, **50**(4), 589–612.

7. McDonald, L. (ed.) (2008) *Florence Nightingale's Suggestions for Thought*, Volume 11 of *The Collected Works of Florence Nightingale*. Waterloo, Ontario: Wilfrid Laurier University Press.

8. Costin, L. B. (1983) *Two Sisters for Social Justice: A Biography of Grace and Edith Abbott*. Urbana: University of Illinois Press.

Michael Dunne-Willows is a Statistical Ambassador for the Royal Statistical Society and a government statistician.

A suspicious sequence?

When a lottery draw produced a straight run of numbers and more than 20 winners, some Twitter users cried "scam". But it does not take foul play to produce such an outcome, says **Michael Dunne-Willows**

On 1 December 2020, the South African lottery drew the sequence 5–6–7–8–9–10, resulting in a jackpot prize being shared by 20 individual tickets. Both the eye-catching sequence of numbers and the unusually large number of winners generated some controversy, with the BBC reporting that some Twitter users had gone so far as to allege a "scam" (bbc.in/30Cgyho).

According to the BBC, "South Africa's National Lotteries Commission (NLC) said it would investigate the draw, which it called unprecedented." However, statistically, the outcome of the draw is not particularly exceptional.

First, let us consider the sequence itself. The draw was made under "Powerball" rules, in which five balls are drawn from a set numbered 1 to 50, followed by a Powerball from a set numbered 1 to 20. In this particular lottery, 8–5–9–7–6 were drawn first, then came the number 10 Powerball, creating a straight-run sequence when arranged in ascending order. A draw like this may seem unlikely – and even suspicious – in isolation. However, we always need to look at the bigger picture.

While it is technically correct that in a given lottery draw, the probability of seeing a specific sequence is 1 in 42 million (bit.ly/2bV4AAf), this figure does not account for the number of lottery draws across the history of the South African lottery. The lottery has been running for approximately 20 years, and there are currently five separate Lotto draws, each taking place every 3 days or so, alongside a stand-alone Daily Lotto draw. If we assume these lottery draws have been introduced gradually over the past 20 years, starting with just one a week, then a back-of-the-envelope calculation shows that there may have been approximately 12,000 draws carried out by the South African lottery since its beginning.

If we assume these 12,000 draws are all Powerball draws (which they are not), the probability of seeing a specific sequence of numbers increases from 1 in 42 million to 1 in 3,500. If we are interested in seeing any one of the 15 possible straight-run sequences of numbers that end with the Powerball as the highest number, the probability of doing so over approximately 12,000 draws is 1 in 235. Still not great odds, but nowhere near our astronomically unlikely starting figure (see "Calculations").

It is important to remember that, in a fair lottery, any sequence of numbers is just as likely as any other. A straight sequence may be eye-catching but, statistically, there is nothing more special about 5–6–7–8–9–10 than there is about the numbers 13–15–24–30–50–9, which happen to be the Powerball numbers that were drawn a week later.

The second part of the story is the unusually large number of winners: 20 people shared the prize for the 1 December draw, whereas jackpots are typically won by individuals or occasionally two or three separate tickets. It is important to remember, though, that people do not necessarily choose their lottery numbers randomly. Some may use significant dates like birthdays or easy-to-remember number sequences such as 2–4–6–8–10–12 or, in this case, 5–6–7–8–9–10. The result of this is that when an eye-catching or popular sequence of numbers is drawn by chance, we are also likely to see a jump in the number of winners sharing the prize.

As a mini experiment, start asking people to pick a number between 1 and 10. You may find that the number 7 comes up far more frequently than others. This is a small-scale example of how people tend not to choose randomly even when they think they are doing so. ■

Calculations

In Powerball rules, there are 42,375,200 possible draw sequences (once sorted into ascending order). This means the probability of a specific sequence occurring in a single draw is

$$\frac{1}{42{,}375{,}200} \approx 0.0000000236$$

Similarly, the probability of a specific sequence having occurred after two draws is

$$\frac{2}{42{,}375{,}200} \approx 0.0000000472$$

This is about 1 in 21 million.

I have estimated the number of total draws in the 20 years of the South African lottery to be approximately 12,000. The probability of seeing a specific sequence occur at least once over this many draws is

$$\frac{12{,}000}{42{,}375{,}200} \approx 0.000283$$

This is about 1 in 3,500.

Now, there are 15 possible draws producing straight-run sequences with the Powerball as the highest number, and we can order them according to their lowest value: 1–2–3–4–5–6, 2–3–4–5–6–7, …, 15–16–17–18–19–20. This means that in each of 12,000 draws, we actually have 15 possible ways to draw one of these straight runs. By multiplying our previous probability by 15 we arrive at the final probability of seeing a straight run (like 5–6–7–8–9–10) over the past 20 years:

$$15 \times \frac{12{,}000}{42{,}375{,}200} \approx 0.00425$$

This is about 1 in 235.

Disclosure statement
The author declares no conflicts of interest.

Brian Tarran

Another round?

In just a few weeks, it'll be exactly one year since I last set foot in the offices of the Royal Statistical Society in London. My Covid-induced isolation started a bit earlier than most, you see, as I developed a cough about a week before the entire country entered lockdown. The cough wasn't anything serious – just something picked up on the commute perhaps, or a bug my children brought home from school. However, then as now, anyone with a new, persistent cough was told to isolate. The only difference back then was that you wouldn't be tested for coronavirus unless you presented at hospital with symptoms. And, as I say, it wasn't that serious a cough.

For much of this past year I have enjoyed my time at home. It was stressful to begin with, trying to balance work and home schooling. And it was impossible to find quiet places to read and write with a house full of people all the time. But that was about the worst of it. And I do realise that my experience could have been so much worse.

Articles by Ron Fricker and Dmitry Kobak this issue, reviewing excess deaths in the United States, Russia and elsewhere (pages 12–19), give some indication of the scale of the tragedy wrought by this pandemic over the past 12 months. For every one of the lives lost to Covid-19, there will be countless others left in mourning. For every case ending in death, there will be many more patients still fighting the disease – some filling wards, needing hospital care; others stuck at home, confined to a room, hoping not to pass on the virus to their loved ones, and hoping that they won't soon need a hospital bed.

Statistics tell the story of the pandemic in their own way. When sending through the final revision of his article in January, Fricker said to me: "The way the numbers increase so substantially from version to version continues to astound me. In early November, the total number of cases in the US was 12 million, which was a huge increase from a month or two before. But in this revision, two months later, it's now 21 million and climbing fast. Incredible."

Words, though, are perhaps better at conveying the fear and helplessness that some have felt as cases rise and rise. Katherine Hoffman's pandemic diary (pages 40–43), covering the first harrowing months of New York City's outbreak last year, offers a flavour of what it was like to be a biostatistician supporting a team of hospital clinicians at that time. It's a gruelling and emotional read that will stay with you.

In fact, the concluding words to Hoffman's article – "I don't know if I can handle another round" – were ringing in my ears at the start of this year, as England entered its third period of national lockdown. I said I was enjoying my time at home, but things feel different now. Worse. The house is full and noisy again now that the schools are closed once more. But my wife now works in our local hospital, which adds to my unease. And people in my family and social circle have tested positive for coronavirus in recent weeks, several of whom have ended up in hospital, a couple of whom have died. I have been hugely fortunate not to have experienced that until now.

One year on, coronavirus continues to take its toll. ■

Statistics tell the story of the pandemic in their own way. Words, though, are perhaps better at conveying the fear and helplessness that some have felt as cases rise and rise

Photo: Elyse Marks Imaging/RSS

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

ASA
AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

Jakayla Toney/Unsplash.com

Covid-19: One year on...

Ron Fricker assesses the impact of the pandemic in the United States by calculating the number of "excess deaths"

The impact of the Covid-19 pandemic is unclear to many people. Some of this is due to the nature and newness of this disease, where our understanding of the SARS-CoV-2 virus and Covid-19 is evolving in real time, and some is due to an "infodemic" of misinformation. For example, at a point when the Covid-19 death toll had exceeded 180,000 in the United States, Donald Trump incorrectly claimed that only 6% of the deaths were actually caused by the virus (cnb.cx/2IZg2L0). Perhaps not surprisingly, a Cornell University study found that Trump "was likely the largest driver of the Covid-19 misinformation 'infodemic'" (nyti.ms/3mvqzLU), but a lack of understanding of the pandemic's impact was and is a worldwide problem (bit.ly/38e7TLj).[1,2]

There are many reasons for this. One is that social media has become a dominant source of information and within that communication ecosystem it is difficult for users to separate truth from fiction (bit.ly/3ph1DJV; bit.ly/3oYTNnK). In the United States, another is that some politicians and broadcast media pundits have spread false or misleading facts, narratives, and explanations to further their self-interests (see, for example, the following *PolitiFact* articles: bit.ly/3r4FUXj; bit.ly/34paqBk; bit.ly/38vDtVh; bit.ly/34rCoMV). In addition, the SARS-CoV-2 virus spreads quickly but subtly and manifests in differential ways (bit.ly/3ak4LAl), making it hard to directly observe cause and effect and thus confounding people's ability to accurately assess their risk of getting Covid-19 (bit.ly/3p86qgD). The cumulative effect is a populace overwhelmed by information yet unsure of what to believe or do.

Illustrating the confusion

Consider the use of Covid-19 case counts as a way to characterise impact. Issues begin with a not uncommon misunderstanding of the definition of a "case". According to the Merriam-Webster dictionary, a medical case is "an instance of disease or injury", but Covid-19 case counts are typically *confirmed* case counts. That is, these counts are instances of the disease that have been substantiated either by a test or medical professional. So, the actual case count must be estimated, a problem that has been exacerbated in the United States which has lagged in testing capability and uniform standards. Furthermore, random testing is necessary in order to accurately estimate the prevalence of Covid-19 (see bit.ly/3nrEbca and bit.ly/3nwvq0l for additional discussion). Yet, in the absence of random testing, the United States has had to rely on less desirable measures, such as the positivity rate, to try to understand the spread of Covid-19 (bit.ly/2WozX9s). Compounding this, the virus affects individuals in about the broadest way possible, meaning some contract the virus and have no symptoms at all and others end up in the hospital or die. Thus, to some, the notion of a case seems either ill-defined or, for asymptomatic cases, incorrectly identified.

As I write this in early January 2021, in the United States the number of confirmed Covid-19 cases currently exceeds 21 million and the number of deaths attributed to Covid-19 exceeds 360,000. While counting Covid-related deaths seems like it might be straightforward, it too has been challenged. When someone dies in the United States, the immediate cause of death, along with up to three underlying conditions that "initiated the events resulting in death", is recorded on a death certificate by a medical professional (bit.ly/3nAvblg). Covid-19 is typically an underlying condition to an immediate cause of death such as pneumonia or acute respiratory distress syndrome (bit.ly/37rSO9U). Unfortunately, some have falsely alleged that medical facilities are incorrectly classifying deaths as Covid-related for financial gain (bit.ly/3gZId99). While not true, for some people it has raised doubts about the accuracy of the number of Covid deaths.

It is thus no wonder that a layperson can become confused about the true impact of the disease. But it is not necessary to appeal to Covid-19 case counts and death counts to get a sense of the magnitude of this pandemic. Instead, let us look at what is referred to as "all-cause" mortality counts, meaning the total number of deaths no matter what the cause.

Mortality statistics

In the United States, death certificates are filed with local health departments which

Dr Ronald D. Fricker Jr is the interim dean of the Virginia Tech College of Science, and a professor in the Virginia Tech Department of Statistics.

then report them to the National Center for Health Statistics (NCHS). As part of the National Vital Statistics System, the NCHS uses this information to tabulate mortality statistics for each state and for the entire country. Once aggregated, the data is publicly available on the Center for Disease Control and Prevention (CDC) website (bit.ly/37xnVB4). According to the CDC's data, there were 2.84 million deaths in the USA in 2018, 2.85 million in 2019, and as of 7 January 2021 an estimated 3.27 million deaths for 2020.

Heart disease is the leading cause of death in the United States, with an annual mortality rate of just over 647,000 deaths per year. The American Cancer Society estimates that in 2020 there were slightly more than 606,500 cancer-related deaths, the second leading cause. At more than 360,000 deaths, Covid is the third leading cause of death in the United States in 2020 as measured by total deaths. To put this in context, in a country of about 330 million people, there were 36,500 motor vehicle fatalities in 2018. In 2017, about 36,000 people died from unintentional falls and about 40,000 from firearms. So, the total number of Covid-related deaths thus far is more than one-half of the number of each of the two leading causes of death. But it is ten times the total annual number of deaths due to firearms, or unintentional falls, or motor vehicle accidents.

Covid-19 is now the leading cause of death in the United States as measured by the number of daily deaths.[3] For example, on 7 January 2021 the US exceeded 4,000 Covid-related deaths in a single day (bit.ly/39LyNuK). At that rate, more people will die of Covid in 10 days than will die from automobile accidents in an entire year.

Calculating excess deaths

A simple comparison of the total number of deaths illustrates the impact of Covid in the USA. With more than 2.8 million deaths in each of 2018 and 2019, 3.27 million deaths in 2020 corresponds to an increase of slightly more than 420,000 deaths compared to the previous two years. That is a 14.8% increase in one year. While the population of the United States has been increasing over the past three years, that increase is less than 1% per year (on average), so the nearly 15% increase in deaths in 2020 is a substantial jump, even accounting for population changes.

To do a more sophisticated analysis requires estimating what 2020 would have been like if the pandemic had never happened. The CDC actually does this using an algorithm based on a Poisson generalised linear model initially developed by Farrington *et al*.[4] and improved upon by Noufaily *et al*.[5] The model is fitted to historical mortality data, where more recent data is adjusted for reporting delays, and it is used to project the weekly mortality under "normal" conditions. (See "The Farrington algorithm", page 15, for a more detailed description.) The difference between the estimated mortality counts from the model and upward deviations in the actual counts then represents the number of excess deaths.

Figure 1 (page 14) shows four years of weekly mortality counts for the United States, compiled by the CDC (bit.ly/37xnVB4) from data submitted by states and the District of Columbia, from the week ending 14 January 2017 to the week ending 26 December 2020. The height of the bars is the weekly mortality count, where prior to 2020 it varied from about 50,000 deaths per week at the lowest point in the summer to about 60,000 deaths per week at the peak in the winter. The black curve is the expected weekly mortality count from the model and the grey curve is the upper bound of the 95% prediction interval for each week, the threshold above which the mortality count is considered to be significantly high.

A number of important aspects of US mortality are evident in the graph. First, a substantial increase in the number of deaths beginning in mid-March 2020 and continuing to the present is unmistakably visible, where the first confirmed Covid-19 case in the United States occurred on 20 January 2020.[6] Since March, mortality in the United States has increased by at least 11.6% compared to the past three years if we conservatively just look at deaths above the threshold, and it could be as much as 14.5% if we look at all deaths above the expected count. It does not take a sophisticated analysis to see that mortality has distinctly and substantially increased during the pandemic when compared to historical trends.

Second, the number of deaths above the threshold since March (the red part of the bars) sums to 328,900 and the number of deaths above the expected counts (the orange and red parts of the bars) sums to 411,702. So, the number of additional deaths since the start of the pandemic in the United States is at least 329,000 and could be as large as 412,000. The number of deaths the CDC currently attributes to Covid-19 as of 6 January 2021 is 356,005 (bit.ly/35uw73j) which is at the low end of this range.

What is behind these additional deaths is not yet completely clear, where it may be that Covid-19 deaths are being undercounted. It also may be that public health measures taken during the pandemic have changed the baseline mortality, perhaps increasing the number of deaths due to other causes. It is likely some of both. For example, a recent study in the *Journal of the American Medical Association* found that 67% of US excess deaths between March and July 2020 documented Covid-19 as a cause of death.[7] But increased mortality from heart disease and two spikes for deaths related to Alzheimer's disease/dementia were also identified during that period. These may be due to delayed medical treatment, perhaps because of the impact of the pandemic on medical facilities, or perhaps because some people did not seek medical treatment to minimise their potential for exposure to Covid-19 (bit.ly/3p6VLCW).

Third, also visible in January 2018 is another period of excess deaths caused by an unusually virulent flu strain that winter. Comparing the two periods plainly shows that the mortality the United States is experiencing in this pandemic is much worse than the flu, even when compared to a year like 2018 in which an estimated 61,000 people died from influenza. Indeed, the weekly number of deaths during this spring and summer have frequently exceeded the peaks in mortality that tend to occur in the winter.

Age and race/ethnicity

Looking a bit deeper into the data, there are differences by age and by race/ethnicity. Figure 2 (page 14) displays the number of deaths by age category for 2015 to 2020. The plot shows that mortality is up in 2020 in all age categories compared to 2015–2019, though note that for those under 25 the numbers were decreasing until 2020. Table 1 (page 14) shows the percentage increase for 2020 compared to 2019. While greater numbers of people are dying in the older age groups, which is natural, somewhat surprisingly given the media coverage,

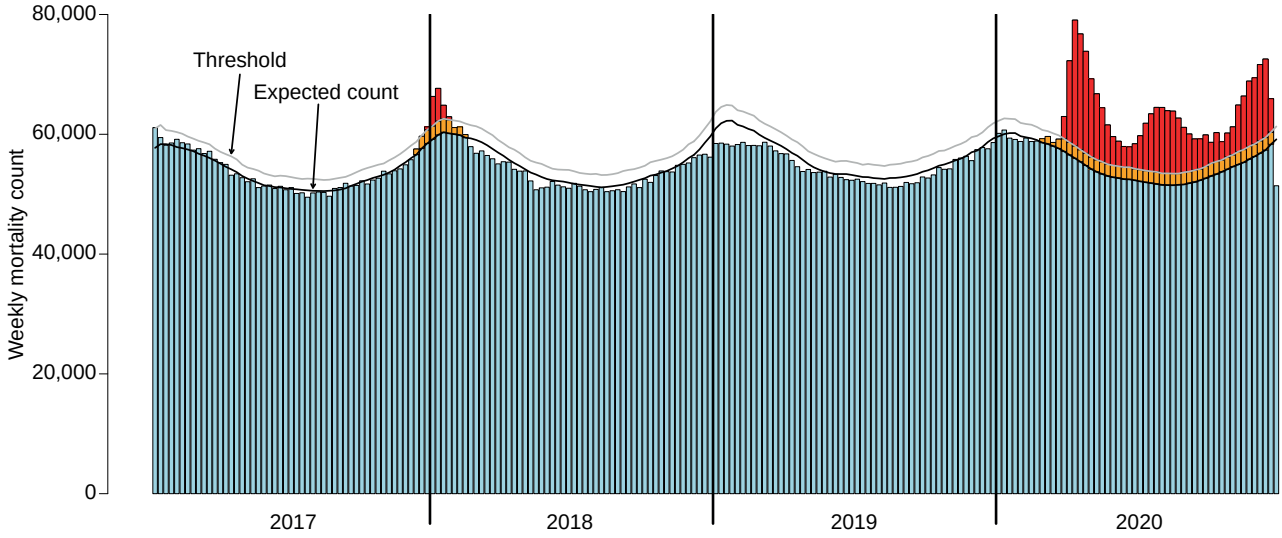


Figure 1: Plot of weekly mortality counts in the United States from the week ending 14 January 2017 to the week ending 26 December 2020. The orange colour denotes the fraction of a weekly count that is above the expected count and below the threshold and the red colour denotes the fraction of a weekly count above the threshold. (The drop in the counts for the last two weeks of 2020 is probably due to data reporting lags exacerbated by the holiday period. Given the current high rate of daily Covid deaths in the United States, these bars will most likely be revised upwards once all the data is reported.)

Covid-19 has had the greatest percentage impact on those in the 25–44 age group.

Figures 3 and 4 show the data by race and ethnicity, where Figure 3 shows increases across all categories, though they are hard to see in some groups because of the differences in magnitude. Figure 4 shows that, while total mortality is higher for non-Hispanic Whites simply due to the population size, the percentage increase across all minority groups from 2019 to 2020 is substantially greater – by a factor of nearly 2 and almost as much as 5 – compared to non-Hispanic Whites.

Conclusions

Using excess deaths as a measure, the impact of the Covid-19 pandemic on the United States should be much clearer. It has resulted in a substantial increase in mortality across all age groups and races/ethnicities, although with a disproportionately greater impact on non-White populations. Also, while much of the reporting and discussion has been on increased mortality among older populations, the greatest percentage increase in mortality is in the 25–44 age group.

Assessing the effects of the pandemic using excess mortality sidesteps the sometimes contentious issues related to whether any particular death was caused by Covid-19. Moreover, excess mortality is useful as an overall measure of the pandemic's impact. For example, as previously mentioned, Woolf *et al.* found increases in heart disease and Alzheimer's disease/dementia-related mortality coincident with the spring surge in Covid-19 cases in the United States.[7] They also found increases in Alzheimer's disease/dementia-related mortality coincident with the summer Covid-19 surge in sunbelt states. These increases may not be directly attributable to Covid-19 but they could be the result of pandemic-related impacts on the health-care system and/or unintended side effects of policies to slow the spread of Covid-19.

That said, the number of deaths attributed to Covid-19 is within the range of excess deaths and, in fact, it is at the lower end of that range. This suggests that the number of deaths currently attributed to Covid may also be an undercount of the actual number of Covid-related deaths.

As a final sobering note, as of 7 January 2021 the Institute for Health Metrics and

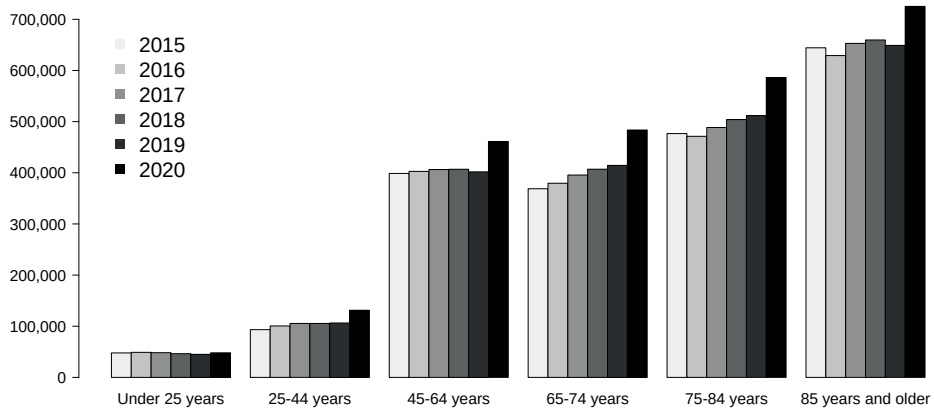


Figure 2: Total US mortality by age category and year.

Table 1: Percentage change in 2020 mortality from 2019 by age category.

Under 25	25–44	45–64	65–74	75–84	85 plus
6.7%	23.4%	14.8%	16.7%	14.5%	11.7%

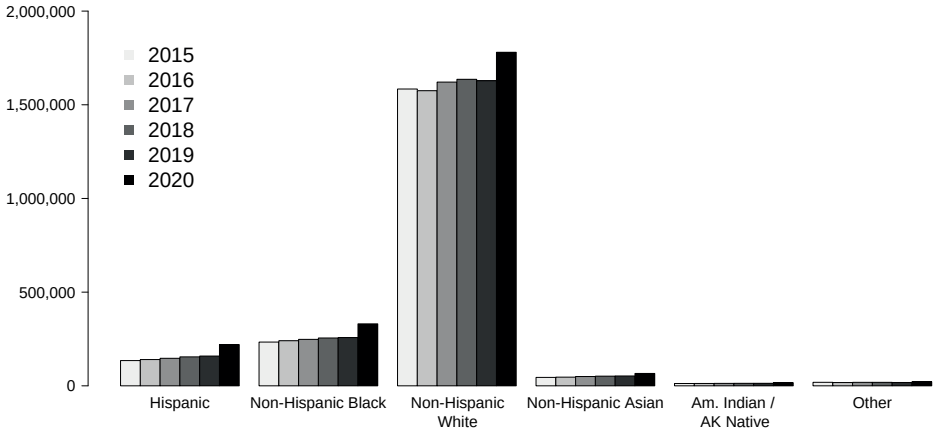


Figure 3: Total US mortality by race/ethnicity and year.

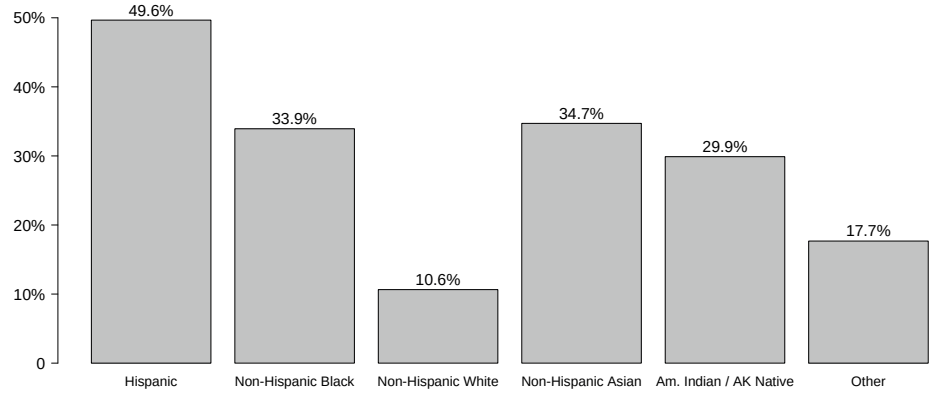


Figure 4: Percentage increase in 2020 mortality compared to the average mortality from 2015 to 2019 by race/ethnicity.

The Farrington algorithm

In health surveillance, algorithms such as Farrington's are used to predict a disease's expected or "normal" state using historical data. Then a substantial increase above what is expected is taken as evidence of a possible outbreak or, in this case, an unusual increase in mortality. Critical in the implementation of any surveillance algorithm is calibrating it to maximise the probability of detecting an increase while constraining the number of false positive signals. These trade off much like Type I and Type II errors do in classical hypothesis testing.[8]

The original Farrington algorithm[4] and its improved version[5] are based on an overdispersed Poisson generalised linear model with spline terms to account for trends such as seasonality in the mortality counts and then to assess deviations in the observed count from the prediction. The algorithm also incorporates logic to address issues of missing data and the presence of a linear trend. The *surveillance* package in R is used by the CDC to implement the Farrington algorithms.[9]

When the Farrington algorithm is used for surveillance, an increase is taken as statistically significant if the observed count exceeds a threshold calculated as a one-sided 95% prediction interval for the next week's mortality count. As employed here, the threshold is used to establish a lower bound on the number of excess deaths, which is the sum of excess counts exceeding the threshold from February to 26 December 2020. In comparison, the sum of the differences between the expected counts predicted by the model and the observed counts is taken as the likely or best estimate of the number of excess deaths.

Evaluation projects more than 567,000 Covid-19 deaths in the United States by 1 April 2021 (bit.ly/3mzsQWa). That puts the number of deaths in the United States due to this pandemic on track to equal or exceed the number of deaths from the 1918 Spanish flu pandemic. While the US population in 1918 was one-third of what it is today, it is no less of a horrifying comparison. Furthermore, it is a tragedy that, in an age when scientists were able to create multiple highly effective Covid vaccines in less than a year, a misinformation "infodemic" has caused the death toll to be far greater than it should have been. ■

Disclosure statement

The author declares no conflicts of interest.

References

1. Editorial (2020) The COVID-19 infodemic. *The Lancet Infectious Diseases*, **20**(8), 875.

2. Galvão, J. (2020) COVID-19: The deadly threat of misinformation. *The Lancet Infectious Diseases*. doi: 10.1016/S1473-3099(20)30721-0.

3. Woolf, S. H., Chapman, D. A. and Lee, J. H. (2020) COVID-19 as the leading cause of death in the United States. *Journal of the American Medical Association*. doi: 10.1001/jama.2020.24865

4. Farrington, C. P., Andrews, N. J., Beale, A. D. and Catchpole, M. A. (1996) A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society, Series A*, **159**, 547–563.

5. Noufaily, A., Enki, D. G., Farrington, P., Garthwaite, P., Andrews, N. and Charlett, A. (2012) An Improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in Medicine*, **32**(7), 1206–1222.

6. Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., *et al.* (2020) First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine*, **382**, 929–936.

7. Woolf, S. H., Chapman, D. A., Sabo, R. T., Weinberger, D. M., Hill, L. and Taylor, D. D. H. (2020) Excess deaths from COVID-19 and other causes, March-July 2020. *Journal of the American Medical Association*, **324**(15), 1562–1564.

8. Fricker, R. D., Jr (2013) *Introduction to Statistical Methods for Biosurveillance, with an Emphasis on Syndromic Surveillance*. Cambridge: Cambridge University Press.

9. Salmon, M., Schumacher, D., and Höhle, M. (2016) Monitoring count time series in R: Aberration detection in public health surveillance. *Journal of Statistical Software*, **70**(10), 1–35.

Excess mortality reveals Covid's true toll in Russia

Data on excess deaths in Russia in 2020 paint a much bleaker picture of the Covid-19 death toll than the official daily updated number, argues **Dmitry Kobak**

In the dashboards that update each day with the cumulative total of Covid-19 cases and deaths across different countries, Russia appears to have fared less poorly than many of its European neighbours and other large nations. At the time of writing (1 January 2021), Russia reports 57,600 deaths (all numbers in the text have been rounded), equivalent to 0.04% of the country's population – far smaller a proportion than in many badly hit countries such as Peru (0.12%), Spain (0.11%), the UK (0.11%), and USA (0.11%). But, in this case, appearances are deceptive.

Besides this daily updated number of Covid-19 deaths, which is included in all international summaries, Russia publishes monthly reports on population dynamics, including the number of Covid-related deaths. These monthly reports appear with a lag of several weeks, and so the last available release at the time of writing covers the month of November. If we add up the monthly summaries from April to November, we see that Russia has had 58,900 deaths from confirmed Covid-19, 12,000 deaths from suspected Covid-19, 11,300 deaths influenced by Covid-19, and 33,800 deaths from unrelated causes in people diagnosed with Covid-19. Based on these numbers, it is 116,000 deaths that should be categorised as Covid-related according to World Health Organization (WHO) guidelines (which recommend counting all deaths in "probable or confirmed" Covid-19 cases, "unless there is a clear alternative cause of death that cannot be related to Covid disease (e.g. trauma)"; bit.ly/34k5P3g). Yet, as of the end of November, Russia's reported death count for Covid-19 in the international dashboards was only 40,500, meaning that the actual number of Covid-related deaths, according to the WHO definition, was almost three times as large.

Russia, of course, is not the only country that has several different ways of counting Covid-related deaths (bit.ly/3p0ILyn). For example, at the time of writing, the UK reports 74,100 deaths through the daily updated numbers, but there are 82,600 deaths "with Covid-19 on the death certificate" (bit.ly/38e5foN). The difference between these two numbers is, however, much smaller than what we see in Russia.

Worse still, Russia's 116,000 Covid-related deaths may not be a reliable estimate of mortality from this disease. It is well understood that the number of confirmed cases cannot be meaningfully compared between countries and even across time because it strongly depends on test availability and testing policy. The same is

Dmitry Kobak is a research scientist at Tübingen University, Germany.

true for the number of deaths, albeit perhaps to a lesser extent: some Covid-19 deaths may go undiagnosed and unreported due to the shortage of tests, or possibly for other reasons as well. An emerging academic consensus is that the most objective way to compare death toll in different countries is via excess mortality.[1–3]

A grim result

Excess mortality refers to the number of deaths from all causes that exceeds the pre-pandemic average. In Russia, the number of deaths from all causes is announced in the same monthly releases in which the different types of Covid-related deaths are reported.

Computing excess mortality in Russia from April to November (see "Estimating excess mortality", page 19, for details) yields a grim result of 264,100 excess deaths (95% interval: [232,000, 296,200]) – see Figure 1. The yearly number of deaths in Russia has been monotonically decreasing over the last decade, and our estimate accounts for that by projecting the linear yearly trend from 2015–2019 into 2020, using that as a baseline to measure the excess mortality; predictive uncertainty gives the standard error. Alternatively, simple subtraction of 2019 deaths from 2020 deaths yields 242,600 excess deaths, while subtraction of the 2017–2019 average yields 230,800 excess deaths. Here we use 264,100 as the most reliable point estimate of excess deaths, meaning that our estimate of excess mortality is 6.5 times as large as the 40,500 deaths reported in the international dashboards during the same period. This estimate of excess mortality from April to November corresponds to 0.18% of the country's population.

Can excess mortality be taken as an estimate of the true Covid-19 mortality in Russia? Some say that it cannot, arguing that lockdown measures that were introduced in most Russian regions at the end of March may have *decreased* the baseline mortality from causes such as violence or traffic accidents, so the true Covid-19 mortality might be greater than the excess mortality. Or perhaps lockdown measures *increased* the baseline mortality due to people's lack of exercise, or through causes related to economic hardship, or because more people died from chronic conditions as routine treatments were postponed or delayed, so the true Covid-19 mortality might be lower than the excess mortality. Either scenario seems plausible. However, comparisons of mortality data between Russia's regions strongly suggest that none of these possible lockdown-related reasons had a noticeable effect on excess mortality.

All mortality figures, including those released monthly, are available for each of the 85 Russian federal regions. By the end of November, excess mortality was above zero in every one of them. Monthly data clearly show the wave of the epidemic as it propagates through the country (see Figure 2a and 2b, page 18). The epidemic started in Moscow, St. Petersburg, and the North Caucasus (Dagestan and Chechnya), where overall mortality increased over the baseline by more than 25% in May. Many neighbouring regions (such as Tver Oblast, Kursk Oblast, and Vladimir Oblast) saw similar increases in mortality in June. Regions in the Ural area (such as Sverdlovsk Oblast, Chelyabinsk Oblast, and Bashkortostan) followed suit in July. Regions in Siberia (such as Novosibirsk Oblast, Kemerovo Oblast, and Altai Krai) first saw increases in mortality of more than 25% in October, and regions in the far east (such as Khabarovsk Krai and Primorsky Krai) saw the same in November.

As stated previously, lockdown measures were introduced in most regions at the same time at the end of March (bit.ly/34lMYFa), and these ran during April and May. This means that many federal regions implemented, and eventually lifted, a strict lockdown regime long before the epidemic reached them. In such regions, overall mortality was mostly unaffected during lockdown: it neither went up nor down, and excess mortality fluctuated around zero. Bashkortostan, for example, showed near zero excess mortality throughout the first half of the year, including the lockdown months of April and May, until a sudden sharp rise in July. This suggests that the lockdown measures on their own, without a local Covid-19 outbreak, did not noticeably influence mortality in Russia.

To assess the contribution that non-Covid causes may have made to excess mortality during an epidemic outbreak, we turn to Moscow and St. Petersburg, two regions with arguably the most reliable reporting of Covid-19 mortality. In both cases, excess mortality is very close to the total Covid-related mortality from the monthly reports (Figure 2c). This suggests that excess mortality can be almost entirely explained by Covid-19 deaths. Mortality from other, non-Covid-related reasons may have also increased – for example, due to the lack of available medical personnel – but the data from these two regions suggest that this had only a minor effect compared to the Covid-related mortality.

If this is the case in Moscow and St. Petersburg, the same should be true in other regions as well. But that is not what one sees in the data (Figure 2c). In most other regions, excess mortality vastly outnumbers the reported Covid-19 mortality, both in the numbers updated each day and in the later monthly reports of Covid-related deaths. The ratios of excess mortality to daily-reported deaths are very different between regions. Moderate, single-digit ratios could possibly be explained by insufficient testing capacity. But many

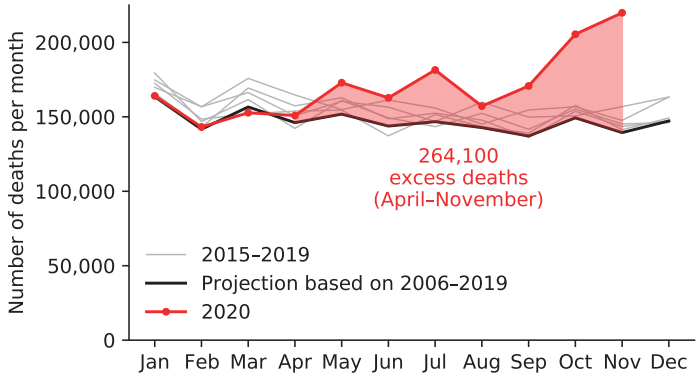


Figure 1: Number of monthly deaths in 2015–2019 (grey lines), projected number of deaths for 2020 based on the previous years (black line), and the actual number of deaths in 2020 (red line). The difference between the black and the red curves gives the excess morality (red shading).

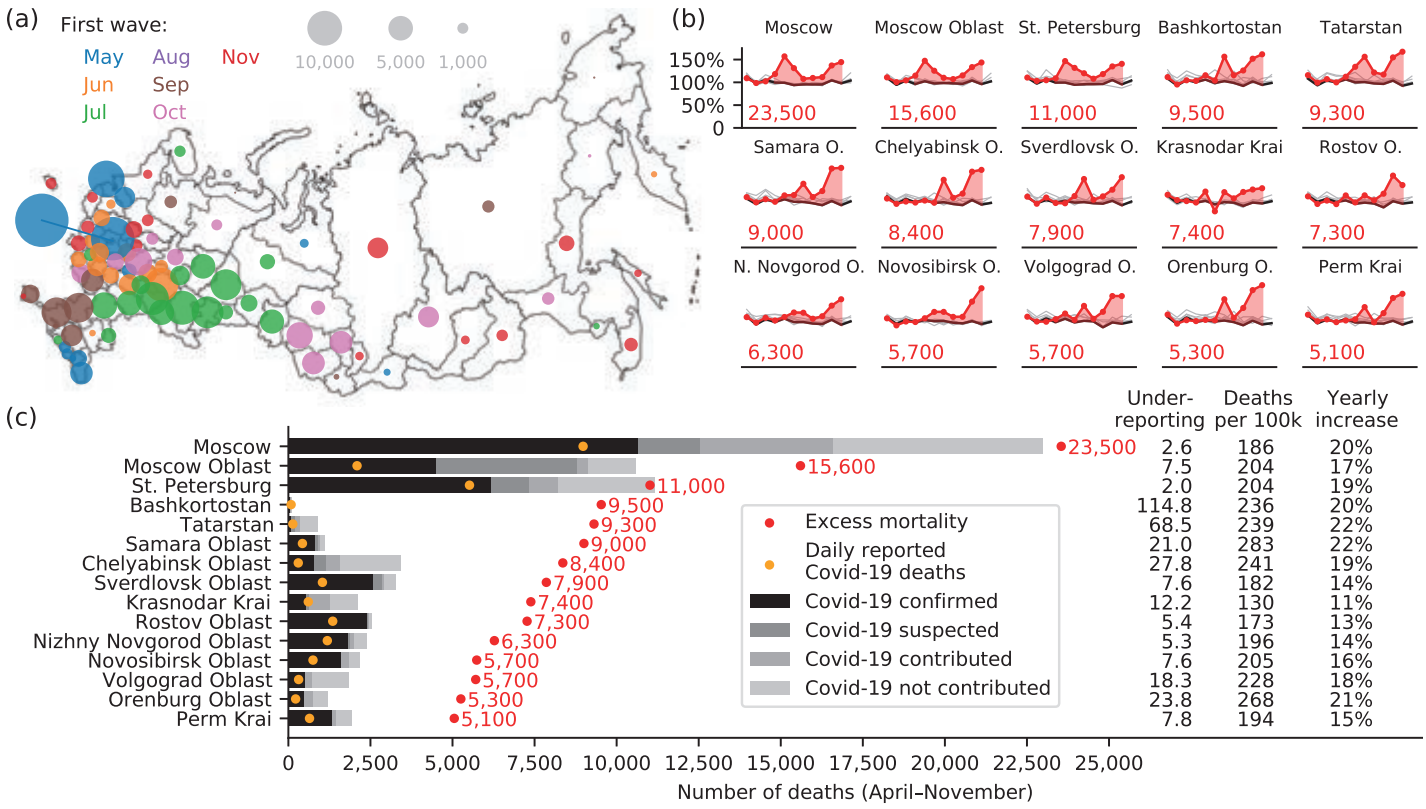


Figure 2: (a) Excess mortality (April to November 2020) in each of the 85 Russian federal regions. The area of each bubble corresponds to the number of excess deaths. The colour corresponds to the month when excess mortality exceeded the baseline by more than 25% for the first time (see legend). See also an animation and further figures at github.com/dkobak/excess-mortality. (b) Time series of excess mortality in the 15 regions with the highest overall excess mortality. (c) Number of Covid-19 deaths (April to November 2020) for the same 15 regions as in (b) measured in six different ways: excess mortality (red dots), daily reported Covid-19 deaths (orange dots), and four categories of monthly reported Covid-related deaths (shades of grey). Columns on the right: Underreporting = ratio of excess mortality to the daily reported deaths; Deaths per 100k = number of excess deaths per 100,000 population; Yearly increase = increase in mortality relative to baseline value for the entire 2020. Among the regions not shown here, the largest yearly increases were in Chechnya (37%) and Dagestan (30%).

regions have excess deaths at more than 20 times the daily reported numbers, up to 30 times in Chechnya, 70 times in Tatarstan, and 110 times in Bashkortostan. This can hardly have a benign explanation and suggests concealment and/or misreporting of Covid-19 deaths. Indeed, there are media reports discussing overflowing hospitals, packed morgues, and deliberate misdiagnosing of Covid-19 as pneumonia (bit.ly/3ar7XKF; cnn.it/3r5u3rH). It may not be a coincidence that Chechnya, Tatarstan and Bashkortostan are among the regions for which there is also statistical evidence of data manipulation in election results.[4,5]

Russia in context

It is not at all unique to Russia that excess mortality is greater than the daily reported numbers of Covid-19 deaths. Several media teams, such as those of the *Financial Times*, *The Economist* and the *New York Times*, have been tracking excess mortality across many countries during the pandemic. Using their latest estimates to compute the ratio of excess deaths to daily reported deaths gives numbers in the range of 0.9 to 3.2 (Figure 3). Note that, ideally, this ratio should be below 1: excess mortality should be smaller than Covid-19 mortality because the latter, according to the WHO guidelines, can include deaths with – but not due to – Covid-19. The ratio of 6.5 that we obtained for Russia is the largest ratio across all countries for which we have data, meaning that the daily reported death count for Russia may well be one of the least reliable indicators of the true epidemiological situation across all countries.

Excess mortality is a lagged indicator, with different lags in different countries. We can, however, use the latest daily report from each country and multiply it by the relevant ratio to estimate, albeit very approximately, what the true Covid-19 death toll might be at that time (Figure 3). For Russia, this gives approximately 380,000 deaths (as of 1 January 2021), corresponding to 0.26% of the population. This is the second highest absolute number of estimated excess deaths in the world after only the USA (over 400,000), and one of the highest numbers per capita: similar to Mexico, Ecuador, Bolivia and Peru (0.25% to 0.28%) and well ahead of all European and North American countries. (Note that these Latin American countries have much younger populations than those of Europe, North America, and Russia, so the same number of deaths per capita in Peru and Russia may indicate substantially higher Covid-19 prevalence in Peru; bit.ly/3b4ojcp.)

Despite all of this, Russian officials have proudly talked about the country's "low"

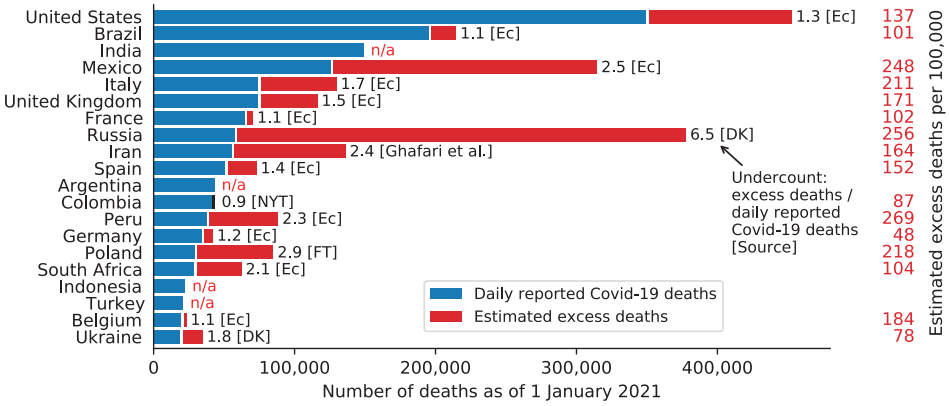


Figure 3: Twenty countries with the highest official daily reported number of Covid-19 deaths as of 1 January 2021. Blue bars are daily reported Covid-19 deaths; red bars are estimated excess deaths as of the same date (daily reported number multiplied by the undercount coefficient). Undercount coefficients are shown near the end of each bar together with the source: Ec = *The Economist* (econ.st/2LKJtlc); NYT = *New York Times* (nyti.ms/3gUMAT9); FT = *Financial Times* (on.ft.com/3bC0Uz3); Ghafari *et al*. (bit.ly/3nu1dPv); DK = this article. The undercount coefficient for Ukraine was computed as described here for Russia, using excess mortality for April–October 2020. Among the countries that are not shown here, the highest undercount coefficient, apart from Russia's, was Bolivia's (3.2). "n/a" denotes countries for which we could not find an estimate of excess mortality. On the right: estimated excess deaths per 100,000 inhabitants.

Estimating excess mortality

The baseline for 2020 excess mortality computations was estimated as follows. Based on the monthly death numbers in 2006–2019, we computed the average monthly number of deaths for each year. This number decreased monotonically apart from a peak in 2010 associated with a summer heat wave that year (go.nature.com/3oILoFj) and a jump upwards in 2015 after Crimea was added to the official Russian numbers. We fitted a linear trend to the 2015–2019 values and extrapolated it to 2020 to obtain the predicted baseline value for 2020 (together with the predictive standard deviation): 147,000 ± 2,000 (Figure 4). Separately, we computed monthly deviations from the average and took the median across all 2006–2019 years to estimate the seasonal variation. Adding the projected average monthly death number to the seasonal variation gives the baseline for 2020. This was done for the entire country as well as for each of the federal regions separately. Our procedure is similar to the approaches used by the *Financial Times*, the *New York Times* and *The Economist*, which also account for linear trends in recent years. The monthly mortality observed in January to March 2020, before the Covid-19 outbreak in Russia, is very close to the predicted baseline (Figure 1), providing support for the baseline estimation. Note that we include Crimea in our analysis despite its contested status because the official statistics in Russia include Crimea.

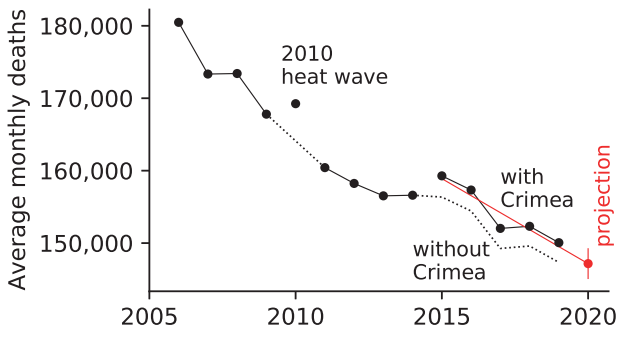


Figure 4: Average monthly deaths in Russia, with and without Crimea. Linear trend and projection for 2020 shown in red.

Covid-19 mortality and the "low" apparent case fatality rate. In June, the president's press secretary said that the "low" death toll in Russia was supposedly due to better health care than in other countries: "Have you ever thought about the possibility of Russia's health care system being more effective?", he asked CNN (cnn.it/2LMJMvR). This rhetorical question could not have been more misleading, as our analysis shows. ■

Postscript

On 28 December 2020, while this article was in preparation, Russian officials suddenly admitted, without any explanation, that most of the excess mortality recorded in Russia between January and November (which was 229,700, they said) was "due to" Covid-19 (bit.ly/394Oa1j). However, all the official data remain unmodified.

Note

Data and code are available at github.com/dkobak/excess-mortality, together with links to data sources, additional figures and animations, and regularly updated data.

The author thanks Maxim Pshenichnikov for discussions and Sergey Shpilkin for scraping and sharing the time series data on regional daily reported Covid-19 deaths.

Disclosure statement

The author declares no conflicts of interest.

References

1. Kontis, V., Bennett, J. E., Rashid, T. *et al.* (2020) Magnitude, demographics and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nature Medicine*, **26**, 1919–1928.

2. Beaney, T., Clarke, J. M., Jain, V., Golestaneh, A. K., Lyons, G., Salman, D., and Majeed, A. (2020) Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? *Journal of the Royal Society of Medicine*, **113**(9), 329–334.

3. Leon, D. A., Shkolnikov, V. M., Smeeth, L., Magnus, P., Pechholdová, M. and Jarvis, C. I. (2020) COVID-19: A need for real-time monitoring of weekly excess deaths. *The Lancet*, **395**(10234), E81.

4. Kobak, D., Shpilkin, S. and Pshenichnikov, M. S. (2016) Integer percentages as electoral falsification fingerprints. *Annals of Applied Statistics*, **10**(1) 54–73.

5. Kobak, D., Shpilkin, S. and Pshenichnikov, M. S. (2016) Statistical fingerprints of electoral fraud? *Significance*, **13**(4), 20–23.

From terrorism to flooding
How vulnerable is your city?

Walter W. Piegorsch, **Rachel R. McCaster** and **Susan L. Cutter** explain how the tools of data science can help quantify the risks and vulnerabilities to hazards in the places where we live and work

When we hear the word "disaster", we often think of events like hurricanes, heatwaves, pandemics, or terrorist attacks. Rarely do we stop to ask how vulnerable our location is to such hazards. Some of us might ponder the question, but the chances are good that any answer we come up with will be somewhat limited in usefulness. Humankind is notoriously poor at judging low-probability, high-consequence events such as pandemics and terrorist attacks, especially when they pertain to adverse, detrimental, "risky" outcomes. Collectively, we are even less aware of any pre-existing vulnerabilities in the places in which we live and work that can either amplify or attenuate the risk of "disaster".

This leads us to a different question: can statistical quantification of the risks and vulnerabilities we face every day become more useful to, and usable by, local residents and decision-makers to understand the dangers and the range of responses communities can muster to address them?

In the absence of quantifiable criteria for assessing a locality's vulnerability to hazardous impacts, risk managers possess only public reactions to subjective, often overly hyped inputs regarding the dangers of potential hazards. Data science can, however, explicitly quantify place-based risk and vulnerability to hazardous impacts.

Quantifying urban vulnerability

In general terms, the vulnerability of places is a function of the social characteristics of the people who live there (social vulnerability) and their susceptibility to harm. Place-based vulnerability is also a function of a community's exposure to damage and loss of function (which we call built-environment vulnerability). Place vulnerability also includes exposure related to physical processes that produce hazardous events such as flooding, hurricanes, or earthquakes, along with their frequency and impact (physical-hazard vulnerability).[1]

For example, markers of social vulnerability might include a locality's per capita income and its percentage of population below the poverty level. Higher values of the former and lower values of the latter afford each household greater potential to prepare for and cope with hazardous events that require increases in household spending; these affect building/structure protection and repair, emergency medical costs, and so on. Another example is a local government's debt-to-revenue ratio. Higher ratios hinder a government's ability to respond to unexpected or sudden hazards: increased debt servicing requirements lower the resources potentially available for response(s) to negative consequences of hazardous events.

For built-environment vulnerability, markers might include a locality's median age of housing units, and its number of mobile homes. Older buildings, if not maintained, suffer greater damage during hazardous events, while large numbers of mobile homes are susceptible to damage in high-wind events such as tornados and hurricanes. The number of hospital beds and modernity of the medical infrastructure is another marker of vulnerability. Smaller and/or older medical institutions cannot respond to or cope with rapid, numerous casualties during a hazardous event. This increases community vulnerability, a relevant issue for contemporary biomedical hazards such as pandemics.

Finally, physical-hazard vulnerability is affected by past hazardous events. A locality's experiences with many past events, especially if they were of diverse forms, indicates the need for more complex protection and mitigation systems. Cumulative experience with hazards – such as frequent flooding – generally drains a community's resources and lowers its resilience to future events, increasing its vulnerability status. The locality's geophysical features – such as peninsulas and islands, extent of shoreline(s), weather/wind patterns – can hinder or prevent rapid evacuation in the period leading up to or immediately following a hazardous event. Also, locations with greater risks of weather-related hazards, such as hurricanes or tornados, carry obvious increased susceptibility to adverse impacts.

By carefully curating and analysing the information these various markers provide, a single new metric can quantify the three broad aspects of place-based vulnerability, focusing on specific undesirable outcomes or fundamental vulnerabilities with which each locality must contend. This single metric is comprised of three different indices:

The *Social Vulnerability Index*[2] (SoVI), first developed in the early 2000s, is designed to summarise socioeconomic and demographic characteristics that interact and influence a community's differential susceptibility to hazardous impacts, along with its overall capacity to prepare for, respond to, and ultimately recover from the event. It is a statistically derived, unitless measure that provides quantitative, comparative values across geographic locations. Larger SoVI scores indicate greater social vulnerability, but these scores have no inherent meaning

Modelling spatial autocorrelation

To incorporate spatial autocorrelation into an analysis of the binary outcome data in our 132-cities database, we chose a construct based on the logistic regression model, called a *centred autologistic model*.[8] Pertinent to our application, spatial autocorrelation was expressly included as a quantitative predictor in the model's construction.[4] For both the urban terrorism vulnerability data and the flood damage data, we calculated the inverse-variance-weighted, place-based vulnerability index, PVI (page 22), for each of the 132 US urban centres and employed it as a single predictor variable, x, in the model. From the consequent model fit, we estimated the autologistic probabilities, $\hat{\pi}(x)$, for each city, in effect ranking them according to their predicted probability of terrorism-related casualties. Table 1 (page 22) lists in order the top 10 cities for the terrorism data according to this arrangement; coincidentally, these also correspond to all those instances where $\hat{\pi}(x) > 0.50$ for this outcome.

We similarly applied the centred autologistic model to our flood vulnerability data. We calculated the model's predicted probabilities of above-median flood-damage claims, $\hat{\pi}(x)$, as a function of x = PVI for each of the 132 cities. Table 2 (page 23) lists the top 10 cities according to this new arrangement. In comparing Tables 1 and 2 we see that five urban centres – Washington, DC, New Orleans, Philadelphia, Norfolk, and Charleston – reside on both lists, exhibiting the greatest probability of adverse outcomes based on both terrorism casualties and flood damage.

Walter W. Piegorsch is a professor in the Department of Mathematics at the University of Arizona, and also serves as the director of statistical research and education at the university's BIO5 Institute. He is the author of *Statistical Data Analytics*, published by John Wiley & Sons (2015).

Rachel R. McCaster is an operational planner for risk management at the South Carolina Emergency Management Division.

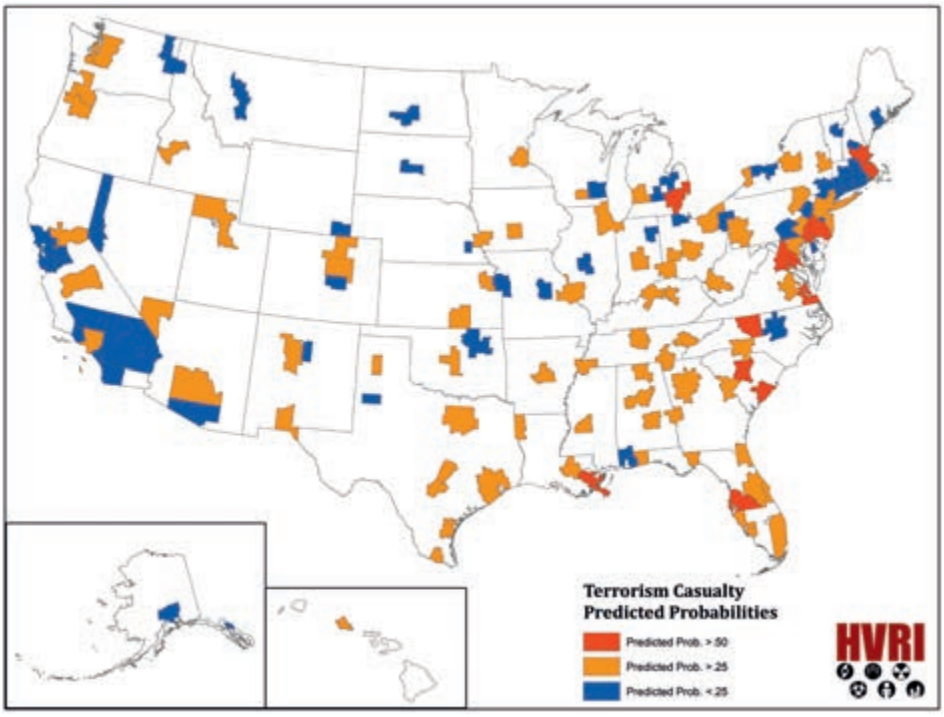


Figure 1: Map of 132 large US metropolitan areas (cities),[3] coded by their predicted probability of terrorism-related casualties.[4] Blue cities, less than 0.25; orange cities, 0.25 to 0.50; red cities, greater than 0.50.

Table 1: Top 10 large US metropolitan areas (cities) with highest autocorrelation-adjusted predicted probabilities of terrorism-related casualties (far-right column). Also included is each city's place-based vulnerability index, PVI, from which the predicted probabilities are calculated.

Metropolitan area ('city')	PVI	$\hat{\pi}$(PVI)
Washington, DC	5.697	0.766
New Orleans, LA	6.838	0.732
Philadelphia, PA	5.456	0.683
Norfolk-Chesapeake-Newport News-Virginia Beach, VA	6.045	0.635
Columbia, SC	4.856	0.587
Tampa-St. Petersburg, FL	4.869	0.579
Greensboro-Winston Salem, NC	4.533	0.562
Charleston, SC	6.262	0.532
Detroit-Warren, MI	3.907	0.521
Boston, MA	4.323	0.514

unless compared to those of other places – generally depicted on a map to visually highlight the comparisons (see sovius.org).

In contrast to SoVI's socioeconomic focus, the *Hazard Vulnerability Index*[1] (HazVI) focuses on geophysical structures that underlie a locality's vulnerability and past hazard experiences; it acts as a surrogate for exposures to and locality-specific involvement with natural events that result in losses within the community. For example, localities in the US state of Nebraska have far lower earthquake risk than those in California, but Nebraska's localities are more prone to tornadoes than California's. HazVI quantifies such geophysical features. It also provides a proxy for potential risk from natural hazards based on the frequency of previous events and the diversity of event types. The latter is important for planning and preparedness purposes: it is much easier to plan for fewer event types and infrequent hazards in both preparedness and response than the reverse.

Expanding on the HazVI metric, the *Built-Environment Vulnerability Index* (BEVI)[1] captures localised vulnerability due to the diversity and type of built-environment infrastructure, such as water and transportation, property values, age of housing, power grid distributions, and support services such as hospitals and fire stations. For instance, large numbers of vulnerable features such as oil and gas lines at risk of leakage or spillage, or bridges vulnerable during earthquakes and flooding, are factors that increase BEVI.

For purposes of summarising a locality's overall vulnerability burden, we combine the SoVI, HazVI, and BEVI indices into a single, place-based vulnerability index, called PVI. Because the indices exhibit different patterns of geographic variability, we construct PVI as a weighted average based on the observed variance of each of the three components: lower weight is given to an index if its variance is large.[3] Higher variability corresponds to lower precision, so this weighted PVI decreases the contribution of a high-variability, low-precision index. We have found the weighted PVI metric to be more effective than a simple, unweighted average for summarising the place-based vulnerability of large urban centres.[3,4] The following examples illustrate its application with two different types of hazard events: urban terrorism and flood damage.

Urban terrorism

In a 2007 paper, we focused on whether or not an urban centre experienced any human casualties (injuries or deaths) from terrorist-related events during a 35-year study period, 1970–2004, for 132 cities in the 50 US states and the District of Columbia.[3] We connected the PVI with data from the Global Terrorism Database (GTD; start.umd.edu/gtd), where a terrorist "event" was quantified via a binary indicator of whether or not any human casualties or deaths were recorded in any terrorism episode during the study period.

Susan L. Cutter is a Carolina Distinguished Professor of Geography and also directs the Hazards & Vulnerability Research Institute at the University of South Carolina. She is a co-editor of *From Disaster to Catastrophe: Emergency Management in the 21st Century*, published by Routledge (2020).

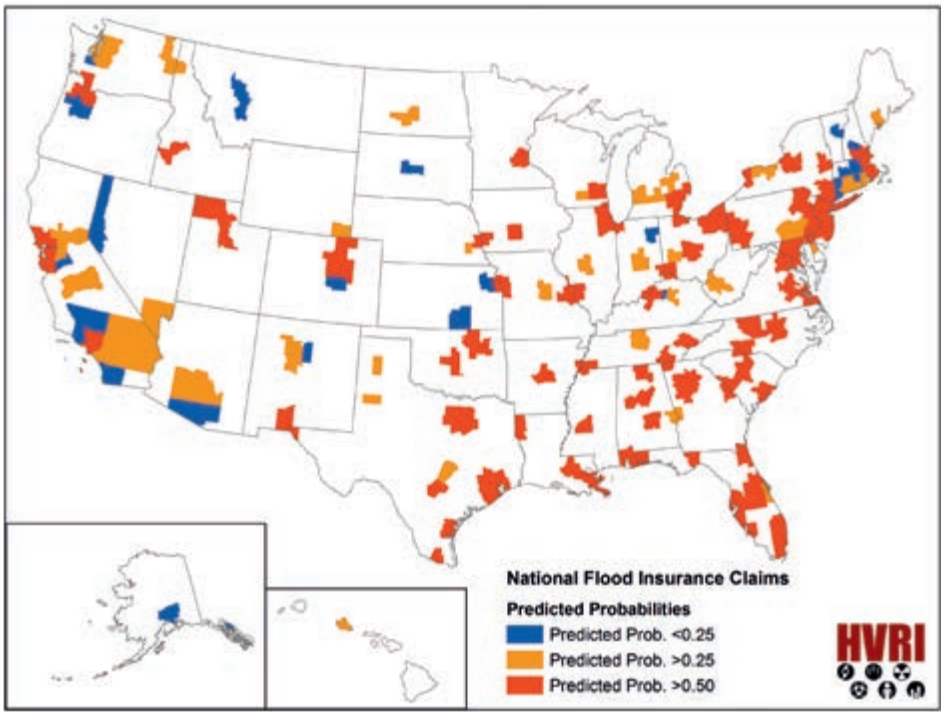


Figure 2: Map of 132 large US metropolitan areas (cities),[3] coded by their predicted probability of flood damage/claims. Blue cities, less than 0.25; orange cities, 0.25 to 0.50; red cities, greater than 0.50.

Table 2: Top 10 large US metropolitan areas (cities) with highest autocorrelation-adjusted predicted probabilities of median-exceeding flood insurance claims (far-right column). Also included is each city's place-based vulnerability index, PVI, from which the predicted probabilities are calculated. Compare to the ordering and values in Table 1.

Metropolitan area ('city')	PVI	$\hat{\pi}$(PVI)
New Orleans, LA	6.838	0.988
Baton Rouge, LA	6.735	0.986
Norfolk-Chesapeake-Newport News-Virginia Beach, VA	6.045	0.978
Charleston, SC	6.262	0.973
New York, NY-Newark, NJ	5.873	0.968
Washington, DC	5.697	0.953
Philadelphia, PA	5.456	0.944
Richmond, VA	5.655	0.937
Houston, TX	5.563	0.933
Boise, ID	5.415	0.918

We did not differentiate the nature, motive, or severity of the terrorist event, because to do so would reduce the number of places with no events. We found that 36 of the 132 cities, or 27%, reported such terrorist-related casualty events during that 35-year time-frame.

Figure 1 maps the full collection of 132 metropolitan areas, colour-coding each locality according to its predicted probability of a terrorist casualty based on its underlying vulnerability, as quantified by the PVI (see Table 1 for a list of the 10 cities with highest PVI-based predicted probabilities). Thus, a city manager studying the map could say that their city has the given predicted probability of a future terrorist attack leading to human casualties, as long as their city registers that particular input PVI. We viewed a predicted probability above 50% as indicative of extreme urban vulnerability (red shading in the figure). Cities with probabilities less than 25% are shaded blue, while those between 25% and 50% are shaded orange. Intriguingly, all cities that exhibit extreme-probability vulnerability are located on or east of the Mississippi River.

In a subsequent article in 2018,[4] we focused on how the PVI predictor was affected by spatial proximity to other cities and found a negative spatial correlation: when a city experiences a terrorist casualty event, an adjacent city would expect *not* to encounter such an event, and *vice versa*. The vulnerability hazardscape depicted in Figure 1 is derived from our 2018 analysis (see "Modelling spatial autocorrelation", page 21).

Notice also in the figure that a number of urban areas outside the highly populous northeast quadrant appear somewhat isolated: especially in the less-populated central and western states, large urban centres are not always adjacent to each other. This is simply a function of our study's focus on only 132 of the largest, most vulnerable, urban centres in the USA and does not hinder the inferences available from the data. (A complete description of the adjacency patterns among these 132 cities appears as supplemental material to our 2018 paper.[4])

Flood damage

One of the most common and most damaging forms of natural hazard is flooding. Flooding causes obvious damage to goods and property, but it is less immediate in showing death and destruction than disasters such as earthquakes and tornados. Yet floods often follow on from hurricanes and severe storms, and they can lead to considerable adverse consequences: damage from hurricanes, storms and flooding accounts for as much as 75% of all US hazard losses in the 50-year period from 1960 to 2009.[5]

An effective data-analytic strategy to quantify flood damage identifies how often insurance claims are submitted by homeowners and businesses affected by severe flood events. The US National Flood Insurance Program (NFIP) provides flood insurance coverage for homeowners and businesses, established by the US Congress in 1968 in response to devastating

Predictive analytics

We can illustrate the predictive capability of the centred autologistic model (see "Modelling spatial autocorrelation", page 21) by applying statistical, "machine" learning techniques to the predictive outcomes. For instance, define a positive prediction for terrorist events as a predictive probability in excess of 50%, $\hat{\pi}$ (PVI) > 0.50, where highly vulnerable cities have the greatest potential to experience a terrorism casualty. Values lower than this represent negative predictions, or simply less risk.

We applied these predictions to the terrorism casualty events observed during the study period (1970–2004) to assess how well the predictions matched actual occurrences. In effect, we trained the predictive model to classify cities as to their potential terrorism status (see Table 3).

A pertinent summary statistic from this 2×2 training table is the *accuracy*, that is, the correct classification rate, also known as the *concordance*. This is the proportion of correct positive and negative predictions: the sum of the two main diagonal counts in the table divided by the total. Here, we find training accuracy equal to 100/132 ≈ 76%, which is above the uninformative, coin-flip baseline of 50%, and is indicative of good predictive power.

For this risk-analytic setting, another pertinent summary statistic is the *precision* – also known as the *positive predictive value* – in the 2×2 table, that is, the correct proportion of positive predictions. The positive predictive value of adverse events for cities concerned about their vulnerability to terrorism casualties is of greater importance than the alternative, negative predictive value of avoiding terrorism casualties. Here, the centred autologistic model's precision equals an encouraging 70%.

The terrorism data occur between 1970–2004, thus we can re-access the GTD and query whether any of these 132 cities experienced terrorism casualties in later years. We compare the predictions from the 1970–2004 training data set with the most recent data from 2005–2018 (called the *test data set* in statistical learning). This produces Table 4.

Now we find test accuracy equal to 86 /132 ≈ 65%, dropping slightly below that from the training data but nonetheless still above 50%. It is not uncommon to see drops in accuracy as the test of a trained classification rule is conducted. Promisingly, precision in the table remains at 70%.

We can also apply a statistical learning analysis on the flood-damage outcomes. The approach is essentially identical: classify a city as positive if its predicted autologistic probability exceeds 50%. Then compare the predicted classifications with those actually observed. This produces Table 5.

Here, test accuracy is 94 /132 ≈ 71%, while precision in the table once again reports as exactly 70%. Both values suggest strong predictive power. Note that more recent data do not currently exist for these flood outcomes that would allow us to construct a test classification analysis.

Table 3: Terrorism casualty training set analysis.

		Observed (1970–2004)		
		Positive adverse event	Negative adverse event	Row totals
Prediction (1970–2004)	Positive adverse event	7	3	10
	Negative adverse event	29	93	122
Column totals		36	96	132

Table 4: Terrorism casualty test set analysis.

		Observed (2005–2018)		
		Positive adverse event	Negative adverse event	Row totals
Prediction (1970–2004)	Positive adverse event	7	3	10
	Negative adverse event	43	79	122
Column totals		50	82	132

Table 5: Flood damage training set analysis.

		Observed (1977–2019)		
		Positive adverse event	Negative adverse event	Row totals
Prediction (1977–2019)	Positive adverse event	49	21	70
	Negative adverse event	17	45	62
Column totals		66	66	132

flood losses from Hurricane Betsy in 1965. Since the NFIP's inception, more than 2 million flood insurance claims have been recorded, producing a substantial source of data on flood events. In 2019, the NFIP released a comprehensive claims data set (bit.ly/34xUw7N). In order to study how urban vulnerability describes and predicts flood losses, we connected claims information spanning the years 1977–2019 in this NFIP data set with the PVI. We employed a binary outcome variable indicating whether a city's number of flood insurance claims was at or above the median for numbers of claims over the entire time period. The resulting analysis provides another opportunity to illustrate place-based patterns of vulnerability: Figure 2 (page 23) maps the full set of 132 cities, again colour-coded according to their predicted probabilities of flood damage/claims. As before, a city manager could refer to this map and say that their city has the given predicted probability of future flood damage, as long as their city registers that particular input PVI.

In Figure 2, 70 out of 132 cities now reside in the extreme-vulnerability category with respect to flooding (shaded in red), and the spatial correlation in the flood data is now positive. In addition, and perhaps not surprisingly, the top 10 cities with probabilities of excess flood insurance claims (Table 2, page 23) now involve localities situated on rivers or shorelines, including New Orleans and Baton Rouge in Louisiana, and Norfolk, Virginia.

Comparing place vulnerability

Comparing Figures 1 and 2, we see that the geographic patterns of predicted probabilities visualise quite differently, with more of the high-flood vulnerability cities appearing in the central and western USA. In particular, the predicted flood-damage probabilities are notably much higher than those in the earlier terrorism-based analysis. In fact, all the top 10 flood-based values are above 90% (Table 2), compared to none of those in the terrorism case: in the latter instance, the highest is Washington, DC at 77% (Table 1). (Here again, these are all values city managers can employ to report their predicted probabilities of future adverse events – flood damage or terrorist casualties – as long as their city registers that particular input PVI.) This suggests that

Melissa Kopka/Bigstock.com

urban vulnerability to flood damage is more extensive across the USA than vulnerability to terrorist casualties. In both cases, we find that the model's predictive capabilities are quite good (see "Predictive analytics", page 24). Overall, the differential patterns in Figures 1 and 2 help illustrate – literally and computationally – how different hazardous outcomes can produce substantively distinct vulnerability hazardscapes.

Responding to risk

These examples illustrate how data-scientific strategies can quantify a location's vulnerability to hazardous events. Our applications to US data on urban vulnerability allow for real knowledge discovery: for example, significant negative spatial correlation was observed for terrorism-based casualties in the database. This may seem counterintuitive at first, but upon reflection it does appear plausible. Perhaps the occurrence of terrorist events in one city tends to increase emergency preparedness and response planning in adjacent cities, leading to fewer terrorism (or at least lowered casualty) events. On the other hand, perhaps putative terrorists ignore nearby cities in order to maximise their desired impact across a wider geographic space. Many other possibilities exist, and understanding the underlying processes that drive terrorist attacks is an open, ongoing research question.[6]

From a larger perspective, the message is simple: it is not difficult to quantify and compare place-based susceptibilities to natural and other hazards; however, to do so, one must think outside the proverbial box and integrate modern place-based vulnerability metrics into the analysis. Indeed, these calculations should be viewed as a foundation from which place-based statistical risk analyses may evolve, as more advanced measures of urban vulnerability – and resilience[7] – are added to the body of work in quantitative risk assessment. ■

Notes and acknowledgement

This research was supported in part by grant no. ES027394 from the US National Institute of Environmental Health Sciences. Sincere thanks are due to Dr Jingyu Liu for background on some of the computational details, and to two anonymous reviewers for constructive inputs on the material.

Disclosure statement

The authors declare no conflicts of interest.

References

1. Borden, K., Schmidtlein, M. C., Emrich, C. T., Piegorsch, W. W. and Cutter, S. L. (2007) Natural hazards vulnerability in U.S. cities. *Journal of Homeland Security and Emergency Management*, **4**(2), 5.

2. Cutter, S. L., Boruff, B. J. and Shirley, W. L. (2003) Social vulnerability to environmental hazards. *Social Science Quarterly*, **84**(2), 242–261.

3. Piegorsch, W. W., Cutter, S. L. and Hardisty, F. (2007) Benchmark analysis for quantifying urban vulnerability to terrorist incidents. *Risk Analysis*, **27**(6), 1411–1425.

4. Liu, J., Piegorsch, W. W., Schissler, A. G. and Cutter, S. L. (2018) Autologistic models for benchmark risk or vulnerability assessment of urban terrorism outcomes. *Journal of the Royal Statistical Society*, *Series A*, **181**(3), 803-823.

5. Gall, M., Borden, K. A., Emrich, C. T. and Cutter, S. L. (2011) The unsustainable trend of natural hazard losses in the United States. *Sustainability*, **3**(11), 2157–2181.

6. Python, A., Illian, J. B., Jones-Todd, C. M. and Blangiardo, M. (2019) The deadly facets of terrorism. *Significance*, **16**(4), 28–31.

7. Cutter, S. L., Ash, K. D. and Emrich, C. T. (2014) The geographies of community disaster resilience. *Global Environmental Change*, **29**, 65–77.

8. Caragea, P. C. and Kaiser, M. S. (2009) Autologistic models with interpretable parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, **14**(3), 281–300.

A needle in (many) haystacks: Using the false alarm rate to sift gravitational waves from noise

Yanyan Zheng, **Marco Cavaglià**, **Ryan Quitzow-James**, and **Kentaro Mogushi** explain how statistics helps scientists spot ripples in the fabric of space and time

What does a weather forecast have in common with the hunt for gravitational waves? Not a lot, you might think. One concerns meteorological conditions here on Earth. The other is about identifying ripples in the fabric of space and time. And yet both activities involve a statistical quantity called the false alarm rate, or FAR for short. You, the reader, also probably use the FAR in your daily life, much more frequently than you might realise. In fact, every time you face a decision that depends on the probability an event may occur, the FAR comes into play.

Suppose you have to travel tomorrow to a faraway place. How do you decide whether to pack an umbrella or a bottle of sunscreen? A good idea would be to check the weather forecast for your destination. If the chance of rain is 90%, you will probably pack an umbrella. Even if there is still a 10% chance that it will be sunny, you will feel pretty confident it will rain. You make this decision by unconsciously estimating how frequently a sunny day may occur at that location when the forecast predicts a 90% chance of rain. If there are only a small number of sunny days when the chance of rain is 90%, you may correctly guess that tomorrow's forecast is reliable – and the smaller the number of sunny days, the more confident you should be. Moreover, when the predicted chance of rain is higher, say 99%, your decision should be more likely to be the right one.

The fraction of sunny days with rain forecasts at or above a given percentage defines the FAR for that prediction level. In technical terms, we say that the FAR is a function of a ranking statistic (in this case, the chance of rain) that defines the likelihood of an experiment's outcome.

Left: Artist's impression of binary black holes about to collide. It is not known if there were any electromagnetic emissions associated with GW190521. Image credit: Mark Myers, ARC Centre of Excellence for Gravitational Wave Discovery (OzGrav).

So, what has this got to do with ripples in space-time? Scientists from the Laser Interferometer Gravitational-Wave Observatory (LIGO) Scientific Collaboration[1] and the Virgo Collaboration[2] use the concept of FAR to determine the likelihood that a signal seen in their detectors is a gravitational wave from a cosmic collision of massive objects in space rather than a terrestrial or instrumental data artefact.

The hunt for gravitational waves

Just as a spoon stirs a cup of coffee, accelerating massive objects stir space and time, generating outward-propagating waves in the geometry of the universe. These waves travel at the speed of light, stretching and compressing the space dimensions as they go.

In the early morning of 14 September 2015, almost a hundred years after Albert Einstein's discovery of general relativity, the twin LIGO detectors in Livingston, Louisiana, and Hanford, Washington, recorded for the first time a gravitational-wave signal from space.[3] The event, called GW150914, originated 1.3 billion years ago from the merger of a pair of stellar-mass black holes into a single, more massive black hole. As the first telescopic observations of Galileo in 1609 marked the beginning of modern astronomy, so the GW150914 detection gave rise to a completely new way of exploring our universe. Since that first detection, LIGO and Virgo scientists have observed tens of these cosmic cataclysmic collisions of black holes and neutron stars,[4] and gravitational-wave astronomy has established itself as a powerful new branch of science to study the dark side of the cosmos.[5] More than 1,500 researchers from over 100 institutions in over 20 countries operate, develop and analyse the data from a world-wide network of gravitational-wave observatories that includes the two LIGO detectors in the USA, the European Virgo detector in Italy, the KAGRA detector in Japan and the GEO600 detector in Germany.[6]

The basic common design of these detectors is that of a modified Michelson interferometer.[7] The LIGO detectors consist of two arms, each 4 kilometres long and orthogonal to one another. They operate by splitting a laser beam at the point where the arms meet (the vertex), with a beam then sent down each arm. Mirrors located at the end of each arm reflect these beams back to the vertex where they interfere and recombine. Time variations in the light of the recombined beam are measured with a photodetector. Figure 1 shows an aerial view of the Louisiana LIGO observatory and a simplified layout of the detector (not to scale).

The lengths of LIGO's arms are tuned relative to each other such that the beams destructively interfere at the vertex, that is, no light reaches the photodetector. When a gravitational wave passes through the interferometer, its arms are rhythmically stretched and compressed. This causes a time-dependent difference in the arm lengths and a variation in the light measured by the photodetector. If a gravitational-wave signal is present in the data, the photodetector output contains information about the amplitude and the phase of the gravitational wave.

The effect of a gravitational wave on the LIGO detector is very small. Typical waves from astrophysical sources warp space-time by a distance less than one ten-thousandth of the diameter of a proton over the length of LIGO's interferometer arms! This amplitude is much smaller than the detector output in the absence of signals, the so-called detector instrumental *background noise*.[8] Therefore, detection of gravitational-wave signals requires extremely sensitive detectors and sophisticated analysis techniques.

A needle in a haystack

Looking for gravitational waves in the detector data is like trying to recognise a song at a very noisy concert. Just as the singer's voice may be covered by the chatter and cheers of the crowd, gravitational-wave signals are typically buried in the detector's background noise. One way to increase the confidence of detecting a signal is to use multiple detectors. If a consistent signal is seen in multiple detectors, there is a higher chance that it comes from space instead of being due to terrestrial noise. For this reason, LIGO and its partners typically implement *time-coincident searches* between different detectors to reject false positives. Since gravitational waves travel at the speed of light, a gravitational-wave signal must be recorded in separate detectors within their light time of flight.

After the detection candidates pass the time-coincident check, they are ranked by a statistic. The ranking statistic used depends on the kind of signal being sought. If the shape of the signal is known from theory, such as in searches for mergers of black holes and neutron stars, the main ranking statistic is the signal-to-noise ratio (SNR).[9] Figure 2 (page 28) shows the theoretical waveform that originates from a binary black hole merger embedded in the detector

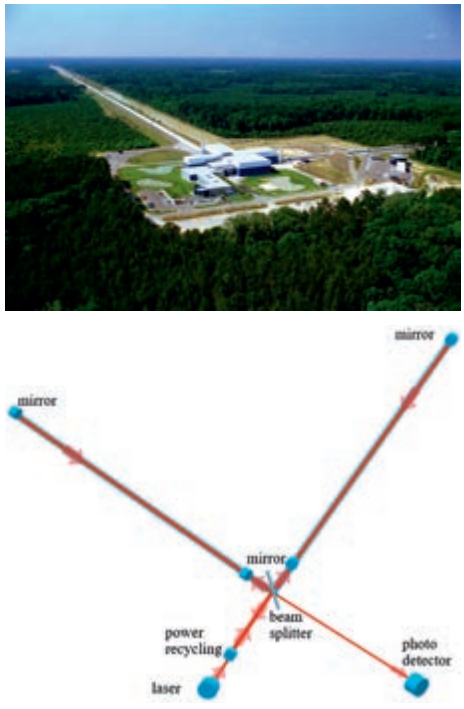


Figure 1: Top: Aerial photograph of the LIGO site in Livingston, Louisiana. Bottom: Simplified diagram of a LIGO detector.

Yanyan Zheng is a physics PhD student at the Missouri University of Science and Technology and a member of the LIGO Scientific Collaboration.

Marco Cavaglià is professor of physics and director of the Institute for Multi-messenger Astrophysics and Cosmology at the Missouri University of Science and Technology. He is a senior member of the LIGO Scientific Collaboration.

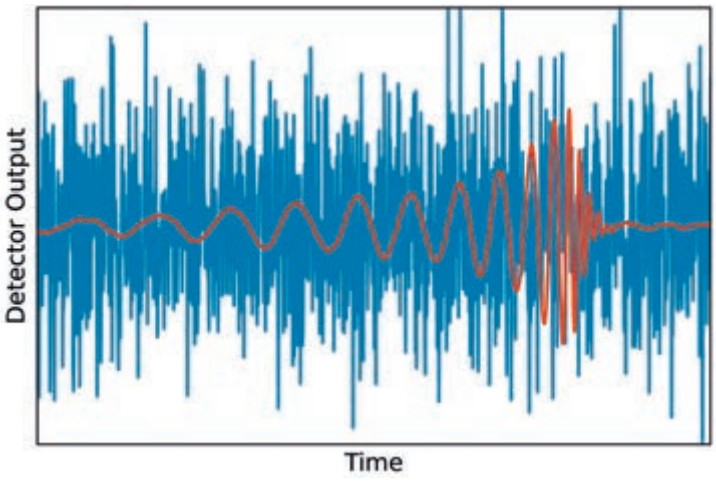


Figure 2: A typical gravitational-wave signal (red) buried in the background noise of the detector (blue). This simulated signal corresponds to the merger of a binary system of two black holes each of mass equal to 40 times the mass of the Sun.

noise. The SNR is roughly proportional to the amplitude of the signal divided by the typical amplitude of the noise. The higher the SNR, the stronger the signal compared to the noise and the more likely it is that the signal can be detected. Thus, the SNR is a good candidate for a ranking statistic to define a FAR. Just as the probability of a sunny day should decrease when the chance of rain becomes higher, the probability that a time-coincident signal in multiple detectors is not a gravitational wave decreases for higher SNR. By setting a threshold on the SNR, we can determine the FAR of the signal candidate and provide a measure of how confident we are that it is real.

Computing the FAR

How do we compute the FAR of a candidate signal with a given SNR? In simple terms, we count the number of background noise events with SNRs equal to or above the SNR of the candidate and then divide by the total analysed time.

The box "False alarm rate and false alarm probability" contains the technical details, but to understand it more intuitively, imagine a gravitational-wave detector as a weather forecaster in a particularly sunny place. Most of the time the forecaster predicts a small chance of rain for the next day, and her prediction turns out to be accurate. However, on some rare days, she gets a strong indication that rain may be on its way and so she predicts a much higher chance of rain. Suppose that tomorrow's predicted chance of rain is 90%. This is equivalent to our SNR in the hunt for gravitational waves. How would we calculate the FAR and the "false alarm" probability (FAP) that tomorrow will nevertheless be sunny despite her predicted 90% chance of rain?

To calculate the FAR and the FAP we need to examine past data. Imagine that in the past 300 sunny days at that location the weather forecaster predicted a chance of rain at or above tomorrow's prediction only three times, and on those three days it was 90%, 95% and 99%. The 300 days constitute our "background" data. To get the FAR of tomorrow's rain forecast, we divide the number of past "false alarms" (3) by the number of background days (300). This gives us a FAR of 3/300 = 0.01 per day (or 3.65 per year), which translates to a 1% FAP of tomorrow being sunny. The FAP would of course be lower (0.3%) if tomorrow's predicted chance of rain were 99% as there was only 1 background event in which the predicted chance of rain was at or above that level. The higher the predicted chance of rain (or SNR, in gravitational-wave detection), the lower the probability of a false forecast (detection).

The more background data we collect, the more accurately we can calculate the FAR. Thus, increasing the amount of background data to analyse is a crucial step of all physical experiments. This is relatively straightforward to do for weather data, for which we have decades of forecasts and actual measurements. When it comes to gravitational waves, the data collected by a detector is limited by the time it has been operating. So gravitational-wave scientists have devised a standard technique, called *time-shifting*,[10] to increase the duration of the background data.

The time-shift technique consists of generating fake coincident data by selecting the data from one detector and shifting the data from all other detectors in time by some arbitrary amount larger than the light time of flight between the detectors. This procedure provides scientists with a set of data that can be used to measure the

False alarm rate and false alarm probability

Mathematically, the FAR of a gravitational-wave signal candidate is defined as

$$\mathrm{FAR} = \frac{N}{T_{\mathrm{BKG}}}$$

where N is the number of detector background noise events with ranking statistic equal to or above that of the candidate event, and T_{BKG} is the total duration of the background data. Under the assumption that the background noise follows the Poisson distribution,

$$P(k) = \frac{e^{-\lambda}\lambda^{k}}{k!}$$

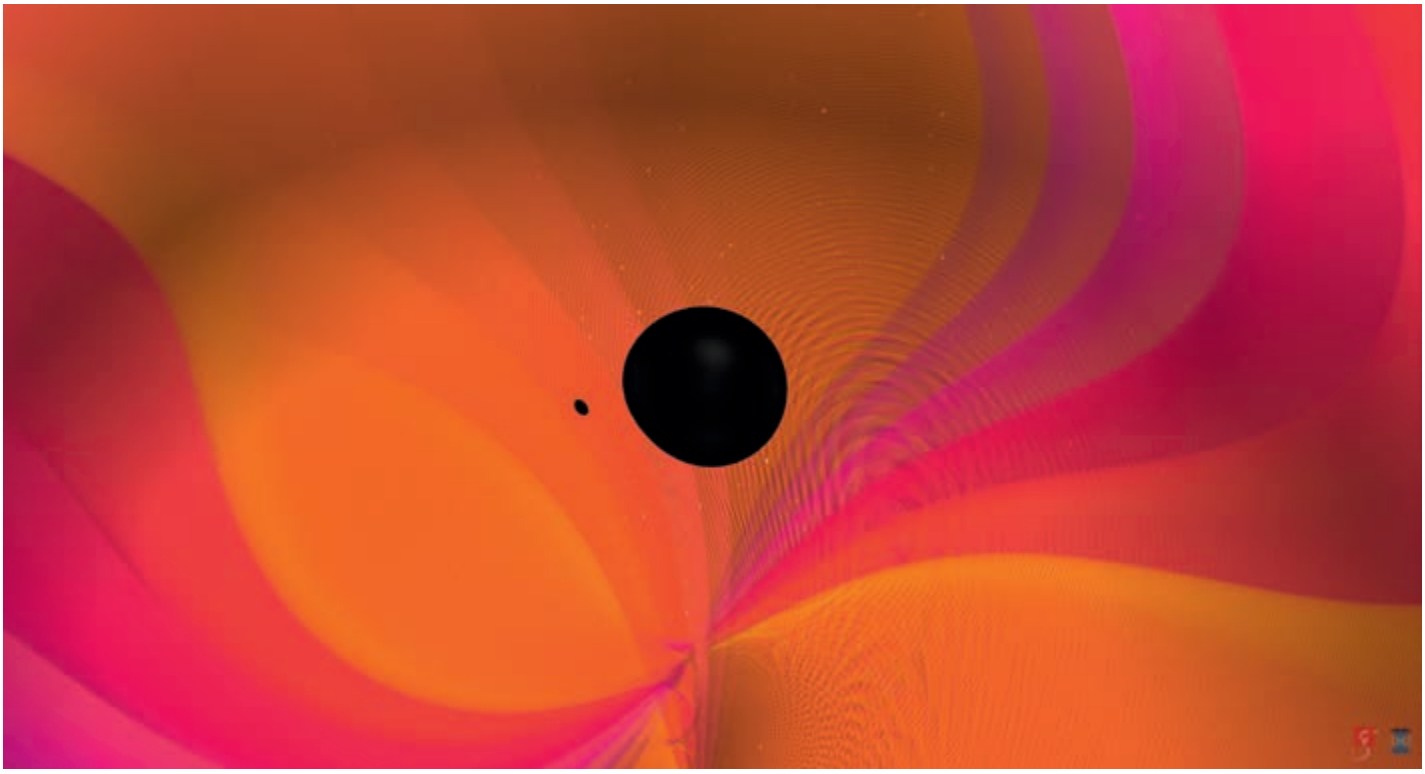
where k is the number of times an event occurs and λ is the average number of events, the (false alarm) probability that a non-astrophysical event with the same ranking statistic of a gravitational-wave candidate occurs at least once in the search time period T_0 is

$$\mathrm{FAP} = 1 - e^{-N(T_0/T_{\mathrm{BKG}})}$$

The FAP provides an alternative way to estimate the significance of a gravitational-wave candidate event. For example, the first gravitational-wave detection, GW150914, has an estimated FAR of less than 1 in 203,000 years, corresponding to a probability of less than 1 in 5,000,000 that the signal was due to terrestrial noise.[3]

Ryan Quitzow-James is a postdoctoral research associate at the Missouri University of Science and Technology. He has been working on gravitational-wave detection and LIGO data analysis since 2009.

Kentaro Mogushi is a PhD student in physics at the Missouri University of Science and Technology. He contributes to the LIGO Scientific Collaboration in various areas.



Above: This image is a still from a video visualization of the coalescence of two black holes that inspiral and merge, emitting gravitational waves. One black hole is 9.2 times more massive than the other and both objects are non-spinning. The high mass-ratio amplifies gravitational wave overtones in the emitted signal. The gravitational-wave signal produced is consistent with the observation made by the LIGO and Virgo gravitational-wave detectors on 14 August 2019 (GW190814).
Credit: N. Fischer, S. Ossokine, H. Pfeiffer, A. Buonanno (Max Planck Institute for Gravitational Physics), Simulating eXtreme Spacetimes (SXS) Collaboration.

number of accidental (false) events which naturally happen because of the background noise. In our weather forecast example, this would be equivalent to looking at forecasts from different meteorologists. If, say, two different weather forecasts predict rain for tomorrow, we could estimate whether this is just an accidental coincidence by shifting all the daily forecasts of one of them by an arbitrary number of days (greater than the typical duration of a storm, say) and measure the likelihood that their forecasts accidentally match.

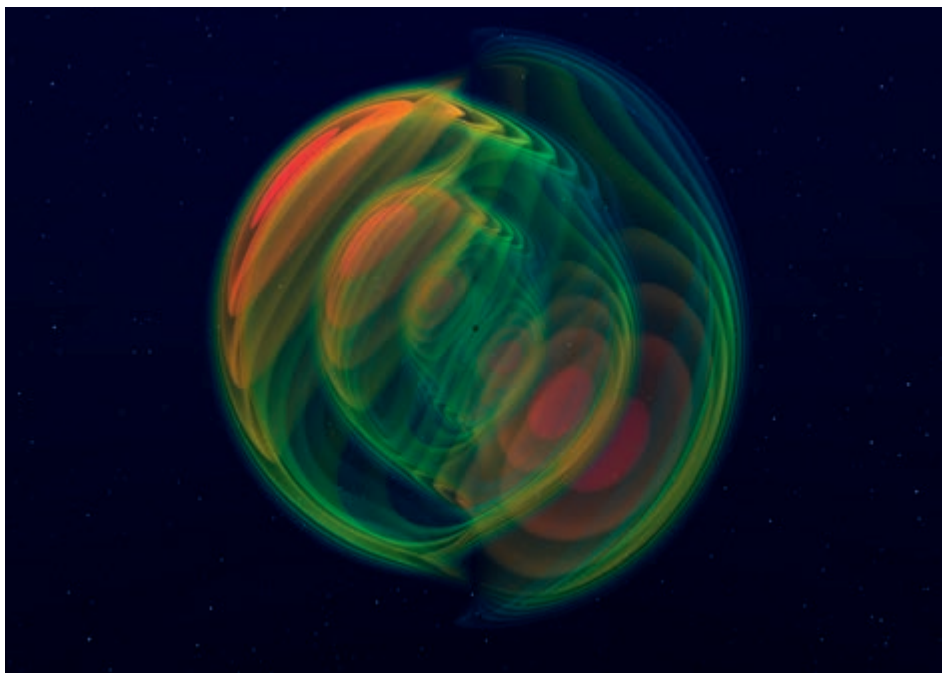


Figure 3: Numerical simulation of gravitational waves emitted by a black-hole binary merger. This event, denoted by GW190412, was discovered by LIGO and Virgo on 12 April 2019.[13] The two merging black holes had masses of about 30 and 8 times the mass of the Sun. The signal has a FAR ranging from less than 1 in 100,000 years to less than 1 in 1,000 years depending on the technique used to recover the signal.
Image credit: N. Fischer, H. Pfeiffer, A. Buonanno (Max Planck Institute for Gravitational Physics), Simulating eXtreme Spacetimes (SXS) Collaboration.

Glossary

Background noise. **Fluctuations in the output of an instrument in the absence of a signal due to instrumental and environmental effects.**

Black hole. **A compact object so dense that even light cannot escape its gravitational pull.**

General relativity**. The theory of gravity proposed by Albert Einstein in 1915. Space and time form a single entity that warps in the presence of matter or energy. The motion of objects is determined by the curvature of space-time.**

Gravitational wave**. The dynamic warping of space-time caused by the accelerated motion of massive objects such as a binary system of orbiting black holes.**

Interference**. Superposition of two or more waves to form a resultant wave. Constructive and destructive interference result from the interaction of coherent waves with the same frequency but different phases.**

Michelson interferometer. **A device that utilises the interference of light waves to perform precise measurements of distance or wavelength.**

Neutron star. **The collapsed core of a massive star. The matter in neutron stars can be more than 10^{14} times denser than water.**

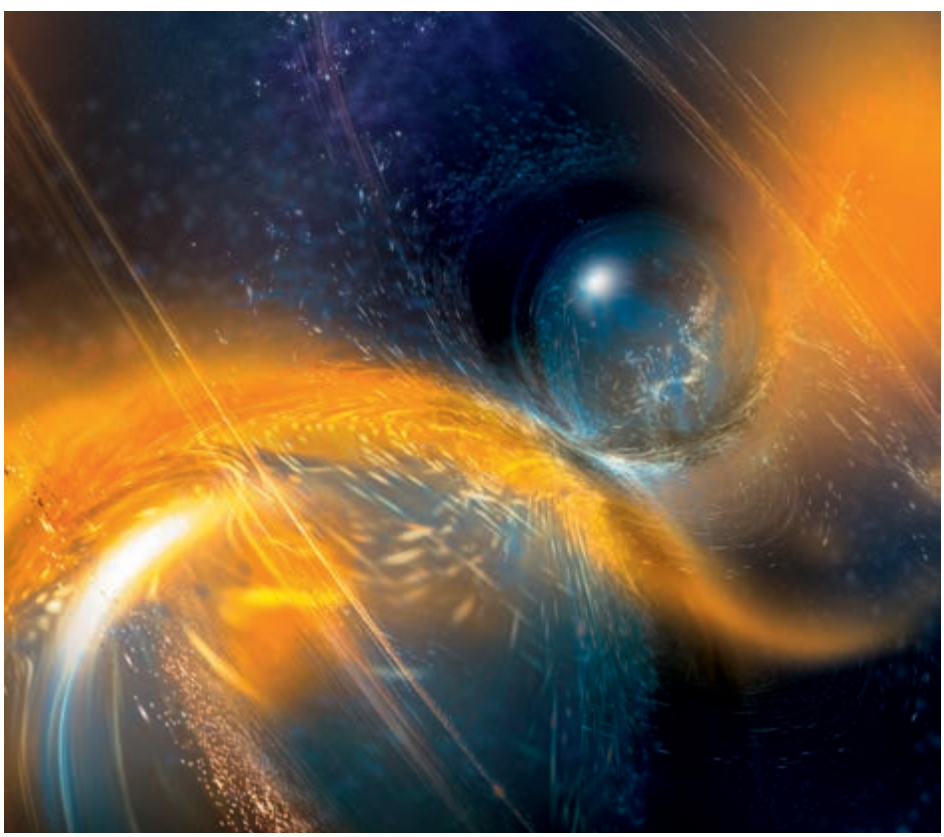
Photodetector. **A sensor that converts light into electrical current.**

Proton. **One of the subatomic particles forming the nucleus of atoms. The estimated radius of a proton is of the order of 10^{-15} metres.**

Sensitivity. **A measure of the smallest signal that a physical instrument is able to detect. The sensitivity of a detector is limited by the background noise.**

Signal-to-noise ratio. **A measure of the level of a signal with respect to the level of background noise.**

Waveform. **A theoretical gravitational-wave signal as predicted by Einstein's theory of general relativity.**

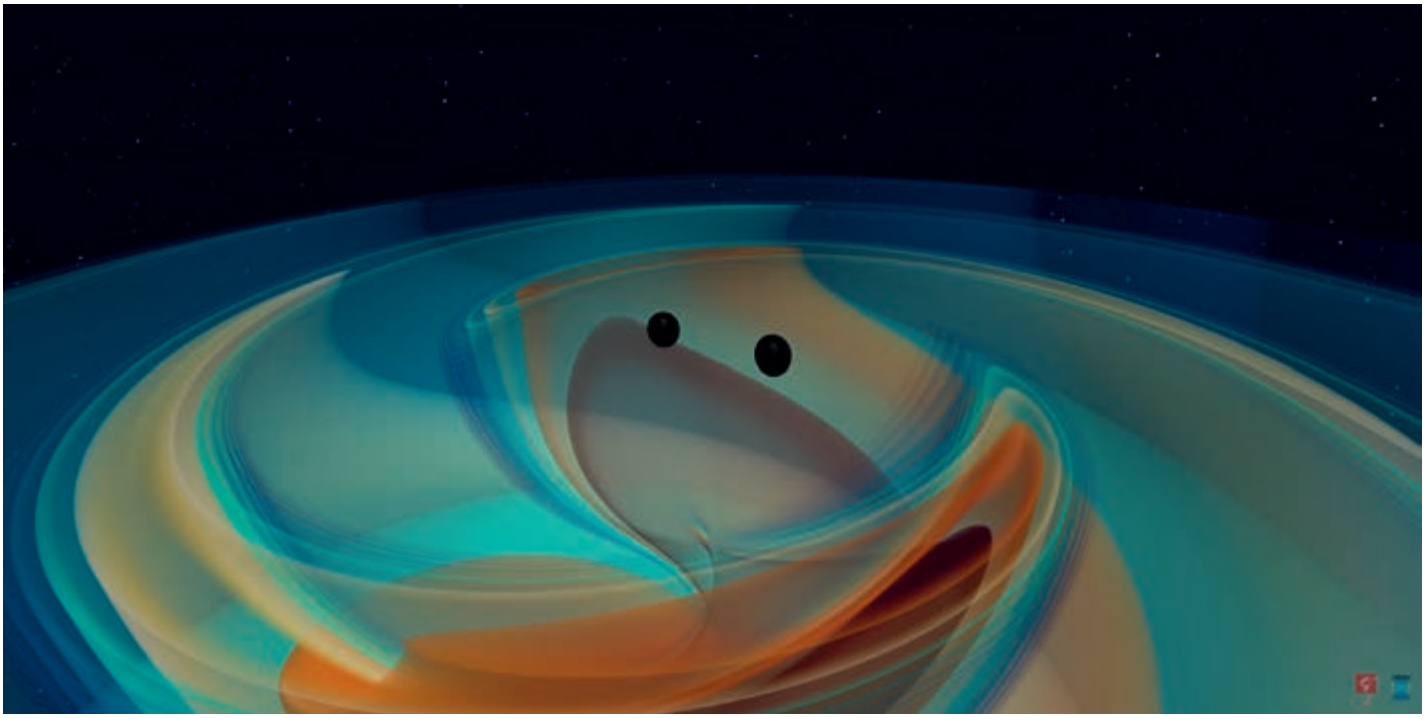


Left: Artist's impression of the binary neutron star merger observed by LIGO Livingston on 25 April 2019 (GW190425). Image credit: National Science Foundation/LIGO/Sonoma State University/A. Simonnet.

In searches for gravitational waves, any candidate signal found in time-shifted data must be caused by random coincidences of instrumental or environmental noise events. Under the assumption that the detector noise does not change too much over time, the background is representative of the time-coincident detector data and can be used to estimate the FAR of a gravitational-wave candidate. The smaller the FAR of an event at a given value of the ranking statistic, the less likely it is that this event is due to the detector's background. In the case of no detections, the FAR allows scientists to set upper limits on the rate of gravitational-wave events. Therefore, the concept of FAR is crucial to investigate gravitational waves from any kind of transient gravitational-wave sources, even from as yet undetected sources such as nearby supernovae.

As we mentioned earlier, the FAR of a gravitational-wave candidate, such as the one depicted in Figure 3 (page 29), depends on the background noise. Thus, it can be made more accurate by improving the detector,[6] mitigating environmental and instrumental noise sources,[11] and improving data analysis algorithms.[12] Many scientists and students in LIGO, Virgo, GEO600 and KAGRA are currently working to make detection techniques increasingly efficient, bring the detectors to design sensitivity, and develop the next generation of gravitational-wave interferometric detectors.

Gravitational-wave scientists also employ a plethora of other statistical tools which could not be covered in this brief article.[8] Statistics played a fundamental role in the detection of gravitational waves and the birth of multi-messenger astrophysics, enabling scientists to look deeply into our universe and understand some of its most fascinating mysteries. As we continue into the future, come rain or shine, this emergent branch of science will continue to rely upon, and benefit from, statistical science. ■



Above: This image is a still from a video of a numerical simulation of a heavy black-hole merger (GW190521). The two black holes inspiral and merge, emitting gravitational waves. The black holes have large and nearly equal masses, with one only 3% more massive than the other. The simulated gravitational wave signal is consistent with the observation made by the LIGO and Virgo gravitational wave detectors on 21 May 2019 (GW190521).
Image credit: N. Fischer, H. Pfeiffer, A. Buonanno (Max Planck Institute for Gravitational Physics), Simulating eXtreme Spacetimes (SXS) Collaboration.

Note

More information on gravitational-wave research can be found by visiting ligo.org, www.virgo-gw.eu, and gwcenter.icrr.u-tokyo.ac.jp/en. Publicly released LIGO data can be found at the Gravitational Wave Open Science Center: gw-openscience.org.

Disclosure statement

The authors are members of the LIGO Scientific Collaboration. LIGO was constructed by the California Institute of Technology and Massachusetts Institute of Technology with funding from the US National Science Foundation under grant PHY-0757058. Computational resources are provided by the LIGO Laboratory and supported by the US National Science Foundation Grants PHY-0757058 and PHY-0823459, as well as resources from the Gravitational Wave Open Science Center (gw-openscience.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. The authors are partially supported by the National Science Foundation under grants PHY-1921006 and PHY-2011334. The authors declare no conflicts of interest or affiliations beyond their academic appointments.

References

1. The LIGO Scientific Collaboration *et al.* (2015) Advanced LIGO. *Classical and Quantum Gravity*, **32**, 074001.
2. Acernese, F. *et al.* (2015) Advanced Virgo: A second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, **32**, 024001.
3. Abbott, B. P. *et al.* (2016) Observation of gravitational waves from a binary black hole merger. *Physical Review Letters*, **116**, 061102.
4. Abbott, B. P. *et al.* (2019) GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Physical Review X*, **9**, 031040.
5. Abbott, B. P. *et al.* (2017) GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Physical Review Letters*, **119**, 161101.
6. Abbott, B. P. *et al.* (2018) Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, **21**, 3.
7. Michelson, A. A. and Morley, E. W. (1887) On the relative motion of the Earth and the luminiferous ether. *American Journal of Science* (third series), **34**(203), 333–345.
8. Abbott, B. P. *et al.* (2020) A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Classical and Quantum Gravity*, **37**, 055002.
9. Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A. and Creighton, J. D. E. (2012) FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Physical Review D*, **85**, 122006.
10. Abbott, B. *et al.* (2005) Search for gravitational waves from galactic and extra-galactic binary neutron stars. *Physical Review D*, **72**, 082001.
11. Abbott, B. P. *et al.* (2018) Effects of data quality vetoes on a search for compact binary coalescences in Advanced LIGO's first observing run. *Classical and Quantum Gravity*, **35**, 065010.
12. Cuoco, E., Powell, J., Cavaglià, M. *et al.* (2020) Enhancing gravitational-wave science with machine learning. Preprint, arXiv:2005.03745 [astro-ph.HE].
13. Abbott, R. *et al.* (2020) GW190412: Observation of a binary-black-hole coalescence with asymmetric masses. *Physical Review D*, **102**, 043015.

"'The scientific method' doesn't leave you"

Timandra Harkness interviews Trevor Phillips, the former head of the UK's Commission for Racial Equality and the Equality and Human Rights Commission, about Covid-19 disparities, ethnic identities and origins, and the often fraught relationship between science and politics

In April 2020, not long after the UK's first coronavirus lockdown began, Trevor Phillips ran what he calls "a quick and dirty exercise" on some Covid-19 data. "It was pretty clear that there was something going on, that there was some relationship between the level of ethnic minority population in a local authority, and the incidence of Covid," he says. "I'm almost embarrassed to say, I just ran the crudest possible regression. Basically, you just line up the local authorities by incidence of Covid death, towards the end of March and April, you just can't escape it."

The pattern that Phillips could not avoid seeing was the disproportionate impact of Covid-19 on ethnic minority populations. It is a pattern we have all become familiar with. Of the people who were critically ill with Covid-19 in England and Wales up to the end of August 2020, just over a third were from non-white ethnic minority backgrounds. The risk of death with Covid-19 is higher for almost every other ethnic group compared to "white British". And according to an October 2020 government report, the rate of death involving Covid-19 was more than three times as high for black males compared to white males during the first wave of infection, and well over twice as high for black females compared to white females (bit.ly/30qi860).

Having spotted a statistical correlation between the hardest-hit local authorities and those with the highest minority ethnic populations, Phillips "went to Number 10" – meaning 10 Downing Street, the official residence of British Prime Minister Boris Johnson – "and said, 'You guys need to pay attention to this.' And to their credit, literally within 48 hours, I was in a meeting with Public Health England."

Number 10 paid attention because, in the corridors of government, Phillips is remembered as the head of the Commission for Racial Equality (CRE) and chair of its successor body, the Equality and Human Rights Commission. Long before that, he studied chemistry at Imperial College, London – which is why you should not be too surprised to hear him talking about regression and data.

"'The scientific method' doesn't leave you," he says. "Most people in public life, when they run across evidence that doesn't fit with their theories about why things are happening, they demand different evidence. And the scientist view is, 'Oh, the evidence doesn't fit my theory, I need to get a better theory'. I'm of that mind."

A case in point: 15 years ago, while heading the CRE, Phillips was presented with data that showed, broadly, that black and Asian school students underperformed their white counterparts. However, as Phillips tells it, more granular data on disparities in educational outcomes then became available, which revealed a more complex picture. Pakistani Muslims and Bangladeshis, for example, were doing comparatively badly, but Indian students were doing well. Phillips suggested the data indicated that the prevailing theory was flawed, that "racist teachers" could not be solely responsible for their students' disadvantage, and the CRE needed to look for other causes.

"I said, 'Look, there's something wrong about our theory, that this is all about racist teachers, because I don't know any teacher who can walk into a classroom and instantly distinguish between the Indian and the Pakistani, and treat the Pakistanis like dirt and treat the Indians like they're princes. I just don't see that happening.' There are some people who have never accepted that and never forgiven me for it. But it's there in front of you. And now [that] we collect much better data, it's just manifest."

Phillips has remained a controversial political figure, and he is unrepentant. "If you want to change life chances for people who are disadvantaged because of their gender or their race, then you've got to know what the problem is, so that you can start to analyse and intervene."

Having spotted a statistical correlation between the hardest-hit local authorities and those with the highest minority ethnic populations, Phillips went to 10 Downing Street and said, "You guys need to pay attention to this"

Understanding origins

When Phillips set about intervening in the Covid-19 story, he hoped to have more to offer than some crude regressions. With Professor Richard Webber he runs a data analysis company, Webber Phillips, that specialises in two things that could be useful in a pandemic.

One is geodemography: knowing the demography of regions that can be as small as individual postcodes. This is familiar data-driven profiling, of the type used for marketing. Richard Webber is perhaps best known for developing the postcode classification system Mosaic for data brokers Experian. Mosaic uses around 800 categories to classify postcodes as "blue collar strivers" or "alpha territory", for example.

Webber Phillips's other specialism is using names to predict people's ethnic and linguistic roots, a system they call Origins. Combining the two, the company's website suggests, can answer questions like, "What parts of London are becoming the new cool areas for Nigerians?" or "How do I find Lithuanians in Burton on Trent?" Such answers may be useful to a local authority wanting to know which streets need Covid leaflets printed in a different language, for example.

Associating names with ethnic and linguistic origins brings its own pitfalls. Phillips is clear that their system could not and should not identify individuals, for example. But even on a population scale, why not just ask people to classify themselves?

Partly, Phillips says, because in practice many people will skip that type of question, making the results too inaccurate to be useful. And partly because, when they do answer, "what people think they are telling you isn't necessarily what you want to know".

He describes an exercise Webber Phillips carried out among people of Turkish Cypriot origin. When asked to self-classify, "something like half of those individuals" ticked the box labelled "white British", because this is how they perceive themselves, says Phillips. But it does not always help to know what people think of themselves, he says, especially when trying to understand discrimination. "People aren't discriminated against because *they* think they're something. They're discriminated against because of what everybody else thinks about them."

Data and profiling can be reductive ways to understand people, because they capture only what is measurable from the outside, not our innermost thoughts and aspirations. But to understand the forces at work on a population scale, affecting each of us in different ways, that outside-in data is sometimes the most important thing.

Webber Phillips applies its Origins analysis for organisations wanting to know how well (or badly) they are doing in representing ethnic minorities at all levels, and for public bodies. "We do a lot of number-crunching for local authorities," says Phillips. "Because the census is 11 years old, it doesn't tell you where people are, and we can detect that very quickly. We're now thinking about moving into the health sphere. And that was provoked by Covid. I think, as we've discovered with Covid,

"Most people in public life, when they run across evidence that doesn't fit with their theories about why things are happening, they demand different evidence. And the scientist view is, 'Oh, the evidence doesn't fit my theory, I need to get a better theory'. I'm of that mind."

Timandra Harkness is a presenter, writer and comedian. Her BBC Radio 4 documentaries include *FutureProofing* and *How to Disagree*, and she is the author of the book *Big Data: Does Size Matter?*

ignorance is lethal. And we just have to get on with understanding the value of data about backgrounds and demography in health in a way that we really shied away from in the past."

Repeated refrains from the UK Prime Minister about "following the science" or being "guided" or "led by the science" are "a terrible mistake", says Phillips

Data gaps

Covid-19 has revealed gaps in available data. Death certificates in England, Wales and Northern Ireland, for example, do not record the ethnicity of the deceased at all, and its recording in Scotland is voluntary. Where organisations do collect data on ethnicity, it is most often self-reported in a few broad categories. Phillips thinks his company could help fill these gaps, as Origins "can segment by 250 different ethnic or linguistic groups, so we can separate Cantonese from Mandarin, Ebo from Yoruba, etc." using only names on records.

But Phillips's past political controversies proved to be an obstacle to his early involvement in this sort of work. "Depressingly, and I still feel furious about this, there was a lot of fuss about us being involved, which meant that we couldn't really get to do any work. I think if we had been able to get on this in April, or early May, we would have been able to predict the outbreaks in Leicester, for example."

Webber Phillips is now working with the Cabinet Office Race Disparity Unit, which produces quarterly reports on Covid Disparities and a plan for action. One of the expert advisors on that team is Dr Raghib Ali, who told me in December 2020 what data from the first wave has revealed about the disproportionate impact of the virus on ethnic minorities in Britain.

"The key finding that we started with was that people from ethnic minority backgrounds did have a higher age-adjusted death rate compared to the white population," he says. "If you just look at the crude figures, then actually the white population has the highest rate because the white population is older, on average. Once you adjust to age, that picture changes completely.

"The second thing is to try to understand what's causing that. And so we look for other risk factors that could explain those differences."

Coronavirus does discriminate by age and sex, it seems. Older people and males are at relatively higher risk. But finding that ethnic minorities are suffering disproportionately could simply reflect that in the UK they tend – on average – to lead different lives than their white counterparts. Ethnicity could be a marker for other risk factors, not a direct cause.

The Office for National Statistics (ONS) has done a lot of work on the various risk factors associated with ethnicity, using modelling and information from the 2011 census (bit.ly/35gE6AZ). "What they found was that where you lived and population density were the biggest factors in explaining increased risk," says Ali. "That explained about half the risk. Occupation was another important risk factor, and also deprivation and household crowding.

"Once you adjusted for all of those, most of the risk in most ethnic groups was explained. There was some residual risk left, particularly for black Africans. The main limitation of the ONS work was that we didn't have data on comorbidities like diabetes and obesity, which are also important risk factors for death."

Another study, Oxford Open Safely, did have data on comorbidities, but not on occupation, so the next task for the Race Disparity Unit report team is to combine data sets to get a multidimensional picture of each patient.

"The UK is the best place in the world to do this kind of analysis," says Ali, "because we're able to link huge amounts of patient data, both primary and secondary care. In other countries it would be very difficult to do this. In the UK we're not going to get the perfect answer, but I think we'll get the best answer, at least with these ethnic groups, anywhere in the world. We'll be able to say, as definitively as possible, how much of it is due to socioeconomic risk factors, how much is due to biological risk factors, and how much is unexplained."

Science and politics

One source of controversy when ethnic disparities in Covid impact first began to be investigated was the question of biological risk factors. Just as males tend to be at higher risk of severe illness and death from Covid than females, was it possible that biological differences contributed to the different impact on different ethnic groups? This speculation received short shrift from Angela Saini, award-winning author of *Superior: The Return of Race Science*, who has warned against "leaping to assumptions of racial difference". She wrote in *Prospect* magazine (bit.ly/2KZYqzU) that: "The job of science here is to account for all external factors until we are left with what can only be biology. The problem is, no researcher has anywhere near the information needed for such exhaustive analysis." (For Dr Raghib Ali's perspective, see "Covid outcomes and ethnicity".)

Trevor Phillips takes a different view. Keen to use all available data to eliminate social factors and see if any unexplained disparities remain, he feels strongly that ruling out biological factors from the start could be a dangerous mistake. "There's a thing called race," he says, "and you cannot pin it to something purely biological. But some population groups are at higher risk than others. And when people say stuff like, 'Oh, race is just a social construct', I'm afraid I see mildly red." His strong feelings on the matter are connected to his own experience. Phillips's family carries the gene for sickle cell anaemia, and as the National Health Service website informs, "most people who carry the sickle cell trait have an African or Caribbean family background" (bit.ly/35fPQnp).

Certainly, we need much more research, much more and better data, and much more knowledge about Covid-19. But that will never tell us what course of action to take, and this troubles Phillips.

"One of the worst parts of the whole debate about science and Covid is that people keep substituting what they want, which is something called 'certainty', with the word 'science'. I worry about it. Scientists are not equipped or elected, or appointed, to make the judgements that politicians have to make.

"There are two big problems here: that science is all about certainty, and that politics is about right and wrong. Politics is not about right and wrong. Politics is always, always, always about what is wrongest and what is a bit less wrong. If there was a right answer, you wouldn't need politics. Everybody would know what it is. What you do in politics is make choices between bad options and worse options."

Repeated refrains from the UK Prime Minister about "following the science" or

Covid outcomes and ethnicity

Dr Raghib Ali is keeping an open mind about what might explain the ethnic disparities in Covid-19 outcomes. However, he is sceptical that biological differences play an important role, and especially sceptical of a genetic explanation, "given that we're looking at very disparate groups". People of black African origin seem to share the highest levels of risk in the UK, but African humans are more genetically diverse than any other set of human populations.

"If you compare two black people, they're more genetically diverse than a white person and a black person," says Ali, "so for some genetic risk to be in black Africans, black Caribbeans, Pakistanis, Bangladeshis, Indians, Filipinos, is all extremely unlikely. The only thing which might be common is that they will have slightly darker skin, so they have slightly lower Vitamin D levels. I'm not convinced it's going to have a major role, but it's possible."

Some studies have suggested associations between genes linked to blood group and more or less serious outcomes from Covid infection, but as Ali points out, in the UK, "the increased risk to ethnic minorities is predominantly because of increased risk of infection, as opposed to increased risk of a poor outcome once infected. So that's why things like population density, occupation, overcrowded housing, are the predominant risk factors. They were much more likely to be infected in the first place. Once they were infected, there's not much evidence they're more likely to die from it."

Research is ongoing into genetic contributions to risk (go.nature.com/3pTUIX7), but as Ali points out, "you can't change people's genes anyway. You need to address what we call modifiable risk factors."

"So, people who are in high-risk occupations should obviously be given appropriate PPE [personal protective equipment] and testing and risk assessment, to see if they have additional risk factors," says Ali. "In terms of where people live, you target your public health campaigning and messaging to those high-risk areas. You can't change overcrowding easily, and you can't change deprivation."

When deciding priorities for vaccination, Ali does not believe ethnic categories are very useful. "The vaccine should be based on your absolute risk, not your relative risk," he says. "So, for example, my relative risk is higher as a British Indian-origin doctor than a white doctor, slightly higher. But people in their sixties, seventies or eighties are at much higher risk, and it should be based on absolute risk of dying once infected." For this, Ali thinks the QCovid Risk Calculator (qcovid.org), developed at the University of Oxford to give individual risk scores, is a useful tool.

"It is true, for example, that if you had a 50-year-old black African male with diabetes or obesity, it might be the same risk as a 60-year-old white female without obesity or diabetes. In both cases, it's not based on ethnicity, it's based on absolute risk. Once you start getting down to the lower age groups, 50 to 60, occupation becomes important there as well. So, if it's bus drivers or security guards or taxi drivers, whatever their ethnic group they should be given priority."

One aspect of the disparity in Covid impact that remains mysterious is what has happened within health care. For example, says Ali, a study of more than 100 deaths of health services staff as of 22 April 2020 found that 94% of doctors who died were non-white (bit.ly/35eTxcO). But only 44% of doctors are non-white. The same study found that non-white nurses and midwives are only 20% of their profession, but they made up 71% of those who died with Covid. "That is quite hard to explain," says Ali, "because particularly the doctors are generally well paid, they're not living in poor areas, overcrowded housing, etc. So, there is further work that needs doing in this area."

being "guided" or "led by the science" are "a terrible mistake", says Phillips. "What he should be saying every day is, 'Look, I've got a judgement to make here. On one judgement, I could do something that makes everybody 100% safe, but I think it means we would be living in huts and hauling carts. Or I could make a judgement that makes us all very rich, but there'll be very few of us because lots would be dead. I'm having to make that judgement every day. The scientists are helping me, and economists are helping me, but as a politician, I'm going to have to make that judgement. And I might be wrong, guys. But you elected me to make that judgement.'"

This seems like wishful thinking. Generally speaking, politicians do not like to admit when they are wrong, or even acknowledge the possibility that their judgements might be flawed, which probably explains why some prefer to seek new evidence than discard a favoured theory. Phillips laments: "We have a political elite, a governing class, a decision-making class, a media class that does not understand the modern world, and how it works. We shouldn't be surprised that they make rubbish decisions." ■

What's the big idea?
"Big Data" and its origins

Against the background of explosive growth in data volume, velocity, and variety, **Francis X. Diebold** investigates the origins of the term "Big Data"

ev/Unsplash.com

The "Big Data" phenomenon, by which I mean the explosive growth in data volume, velocity, and variety, is at the heart of modern science (and is similarly central to modern business). Indeed, the necessity of grappling with Big Data, and the desirability of unlocking the information hidden within, is now a key theme in all the sciences – arguably the key scientific theme of our times.

Parts of my field of econometrics, to take a tiny example, are working furiously to develop methods for learning from the massive amount of tick-by-tick financial market data now available.[1] In response to a question like "How big is your data set?" in a financial econometric context, an answer like "90 observations on each of 10 variables" would have been common 50 years ago, but now it is comically quaint. A modern answer is likely to be a file size rather than an observation count, and it is more likely to be 200 gigabytes (GB) than the 5 kilobytes (say) of 50 years ago. Moreover, someone reading this in 20 years will surely have a good laugh at my implicit assertion that a 200 GB data set is large. In other disciplines such as physics, 200 GB is already small. The Large Hadron Collider experiments that led to the discovery of the Higgs boson, for example, produce a petabyte (10^{15} bytes) of data *per second*.

My interest in the origins of the term "Big Data" was piqued in 2012 when Marco Pospiech, at the time a PhD student studying the Big Data phenomenon at the Technical University of Freiberg, informed me in private correspondence that he had traced the use of the term (in the modern sense) to my paper, "'Big Data' Dynamic Factor Models for Macroeconomic Measurement and Forecasting", presented at the Eighth World Congress of the Econometric Society in Seattle in August 2000.[2,3]

Intrigued, I did a bit more digging. And a deeper investigation reveals that the situation is more nuanced than it first appears.

Big Data and me

I stumbled on the term "Big Data" innocently enough, via discussion of two papers that were presented alongside mine at the Eighth World Congress of the Econometric Society.[4,5] These papers took a new approach to macroeconometric dynamic factor models (DFMs): simple statistical models in which variation in a potentially large set of serially correlated observed variables is driven in part by their dependence on a small set of underlying serially correlated latent variables, or "factors". DFMs are popular in dynamic economic contexts, in which observed variables often move closely together.[6]

Early vintage DFMs included just a few variables, because parsimony was essential for tractability of numerical likelihood optimisation.[7] The new work, in contrast, showed how DFMs could be estimated using statistical principal components in conjunction with least squares regression, thereby dispensing with numerical optimisation and opening the field to analysis of much larger data sets while nevertheless retaining a likelihood-based approach.

My discussion had two overarching goals. First, I wanted to contrast the old and new macroeconometric DFM environments. Second, I wanted to emphasise that the driver of the new macroeconometric DFM developments matched the driver of many other recent scientific developments: explosive growth in available data. To that end, I wanted a concise term that conjured a stark image. I settled on the term "Big Data", which seemed apt and resonant and intriguingly Orwellian (especially when capitalised), and which helped to promote both goals.

Murky origins

My paper seems to have been the first academic reference to Big Data in a title or abstract in the statistics, econometrics, or additional *x*-metrics (insert your favourite *x*) literatures. Moving backwards in time from there, things get murkier. It seems academics were aware of the emerging phenomenon but not the term.[8] Conversely, a few pre-2000 references, both academic and non-academic, used the term but were not thoroughly aware of the phenomenon.

On the academic side, for example, Tilly (1984) mentions Big Data, but this article is not about the Big Data phenomenon and demonstrates no awareness of it; rather, it is a discourse on whether statistical data analyses are of value to historians.[9] On the non-academic side, the margin comments of a computer program posted to a newsgroup in 1987 mention a programming technique called "small code, big data" (note the absence of capitalisation; bit.ly/3r3sdIJ).

Next, Eric Larson provides an early popular-press mention in a 1989 *Washington Post* article about firms that assemble and sell lists to junk-mailers. He notes in passing that: "The keepers of Big Data" – now capitalised – "say they do it for the consumer's benefit."[10] Finally, a 1996 PR Newswire, Inc. release mentions network technology "for CPU clustering and Big Data applications".

There is, however, some pre-2000 activity that is spot-on. First, on the industry side, "Big Data" the term, coupled with awareness of "Big Data" the phenomenon, was clearly percolating at Silicon Graphics (SGI) in the mid-1990s. John Mashey, retired former chief scientist at SGI, produced a 1998 SGI slide deck entitled "Big Data and the Next Wave of InfraStress", which clearly demonstrates this awareness (bit.ly/34lGqWS). Relatedly, SGI ran magazine ads that featured the term "Big Data" in *Black Enterprise* in 1996, several times in *Info World* starting November 1997, and several times in *CIO* starting February 1998. Clearly, Mashey and the SGI community were onto Big Data early, using it both as a unifying theme for technical seminars and as an advertising hook.

Francis X. Diebold is the Paul F. Miller, Jr. and E. Warren Shafer Miller professor of social sciences, and professor of economics, finance and statistics at the University of Pennsylvania.

Second, on the academic side, in the context of computer graphics, Cox and Ellsworth (1997) describe "an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk", which they call "the problem of big data".[11] In addition, Weiss and Indurkhya (1998) note that "very large collections of data ... are now being compiled into centralized data warehouses, allowing analysts to make use of powerful methods to examine data more comprehensively. In theory, 'Big Data' can lead to much stronger conclusions for data-mining applications, but in practice many difficulties arise."[12]

Finally, arriving on the scene later but also going beyond previous work in compelling ways, Laney (2001) highlighted the "three Vs" of Big Data (volume, variety and velocity) in an unpublished 2001 research note at META Group.[13] Laney's note is clearly relevant, producing a substantially enriched conceptualisation of the Big Data phenomenon. In short, if Laney arrived slightly late, he nevertheless brought more to the table.

The rest, as they say, is history. As the synopsis for a recent BBC Radio 4 podcast (bbc.in/3agQcgQ) puts it:

> In 2012, Big Data entered the mainstream when it was discussed at the World Economic Forum in Davos. In March that year, the American government provided $200 million in research programs for Big Data computing. Soon afterward, the term was included in the Oxford English Dictionary for the first time.

A discipline, and a triumph?

Big Data is arguably now not only a phenomenon and a well-known term, but also a discipline. To some, that might sound like marketing fluff, and it can be hard to resist smirking when told that major firms are rushing to create new executive titles like "Vice President for Big Data".[14] But the phenomenon behind the term is real, so it may be natural for a corresponding new business discipline to emerge, whatever its executive titles.

Business discipline or not, it is still not obvious that Big Data constitutes a new *scientific* discipline. Sceptics will argue that traditional areas such as statistics and computer science are perfectly capable of confronting the new phenomenon, so that Big Data is not a new discipline, but rather just a box drawn around some traditional ones. It is hard not to notice, however, that the whole of Big Data seems greater than the sum of its parts. That is, by drawing on perspectives from a variety of traditional disciplines, Big Data is not merely taking us to (bigger) traditional places; rather, it is taking us to very new places, unimaginable only a short time ago. In a landscape littered with failed attempts at interdisciplinary collaboration, Big Data is an interdisciplinary triumph.

As always, however, there is a flip side. Big Data pitfalls may lurk, for example, in the emergence of continuous surveillance facilitated by advances in real-time massive data capture, storage, and analysis. As George Orwell wrote in his famously prescient novel, *Nineteen Eighty-Four*:

> Always eyes watching you and the voice enveloping you. Asleep or awake, indoors or outdoors, in the bath or bed – no escape. Nothing was your own except the few cubic centimetres in your skull.[15]

Time will reveal how Big Data opportunities and pitfalls evolve and interact. ■

Disclosure statement

The author declares no conflicts of interest.

Acknowledgements

For helpful comments, I thank (without implicating anyone in any way): Larry Brown, David Cannadine, Xu Cheng, Tom Coupe, Flavio Cunha, Susan Diebold, Melissa Fitzgerald, Dean Foster, Michael Halperin, Steve Lohr, John Mashey, Tom Nickolas, Lauris Olson, Mallesh Pai, Marco Pospiech, Brian Tarran, Frank Schorfheide, Minchul Shin, Mike Steele, and Stephen Stigler.

References

1. Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2013) Financial risk measurement for financial risk management. In G. Constantinides, M. Harris and R. Stulz (eds), *Handbook of the Economics of Finance*, Vol. 2B (pp. 1127–1220). Amsterdam: North-Holland.

2. Diebold, F. X. (2000) Big Data dynamic factor models for macroeconomic measurement and forecasting. Paper presented to Eighth World Congress of the Econometric Society, Seattle, August. bit.ly/3nwXMI5

3. Diebold, F. X. (2003) "Big Data" dynamic factor models for macroeconomic measurement and forecasting: A discussion of the papers by Reichlin and Watson. In M. Dewatripont, L. P. Hansen and S. Turnovsky (eds), *Advances in Economics and Econometrics: Theory and Applications. Eighth World Congress of the Econometric Society* (pp. 115–122). Cambridge: Cambridge University Press.

4. Reichlin, L. (2003) Factor models in large cross sections of time series. In M. Dewatripont, L. P. Hansen and S. Turnovsky (eds), *Advances in Economics and Econometrics: Theory and Applications. Eighth World Congress of the Econometric Society* (pp. 47–86). Cambridge: Cambridge University Press.

5. Watson, M. W. (2003) Macroeconomic forecasting using many predictors. In M. Dewatripont, L. P. Hansen and S. Turnovsky (eds), *Advances in Economics and Econometrics: Theory and Applications. Eighth World Congress of the Econometric Society* (pp. 87–115). Cambridge: Cambridge University Press.

6. Stock, J. H. and Watson, M. W. (2011) Dynamic factor models. In M. P. Clements and D. F. Hendry (eds), *The Oxford Handbook of Economic Forecasting*. New York: Oxford University Press.

7. Geweke, J. F. (1977) The dynamic factor analysis of economic time series models. In D. Aigner and A. Goldberger (eds), *Latent Variables in Socio-economic Models* (pp. 365–383). Amsterdam: North-Holland.

8. National Research Council (1996) *Massive Data Sets: Proceedings of a Workshop, Committee on Applied and Theoretical Statistics*. Washington, DC: National Academies Press. bit.ly/3bVWGCP

9. Tilly, C. (1984) The old new social history and the new old social history. *Review (Fernand Braudel Center)*, 7, 363–406.

10. Larson, E. (1989) They're making a list: Data companies and the pigeonholing of America. *Washington Post*, 27 July.

11. Cox, M. and Ellsworth, D. (1997) Application-controlled demand paging for out-of-core visualization. In *Proceedings of the 8th Conference on Visualization* (pp. 235–244). Washington, DC: IEEE Computer Society Press.

12. Weiss, S. M. and Indurkhya, N. (1998) *Predictive Data Mining: A Practical Guide*. San Francisco: Morgan Kaufmann Publishers.

13. Laney, D. (2001) 3-D data management: Controlling data volume, velocity and variety. META Group Research Note, File 949, 6 February.

14. Lohr, S. (2012) How Big Data became so big. *New York Times*, 11 August. nyti.ms/3oZ1lXL

15. Orwell, G. (1949) *Nineteen Eighty-Four*. London: Secker & Warburg.

William Isaac

"It is almost a necessity for researchers to collaborate and interact with people from other disciplines"

When I started out in my career, I was very interested in polling and analysis of political data. I expected that I was going to end up doing data science work or statistical analysis in an academic setting, focusing on political data. But it just so happens that I started working on the subject of predictive policing and algorithms at a time when this became a really pressing public issue, and that pushed my career in a different direction.

My undergraduate and master's degrees were in political science and public policy, respectively, but I always had an interest in quantitative analysis. While I was at George Mason University pursuing my master's, I was also working at a Washington, DC, think tank, one specialising in environmental issues. We did a lot of work for the Environmental Protection Agency and the Department of Energy, a lot of quantitative analyses on regulatory matters. But then I decided I wanted to get my doctorate, so I went to Michigan State University to do a PhD in political science.

It was while I was in graduate school that I started working with the Human Rights Data Analysis Group (HRDAG), and one of the first topics I worked on was predictive policing systems. At the time, these systems were not so well known in the public discourse as they are now. A colleague at HRDAG, Kristian Lum,[1] and I came up with a project to evaluate a predictive policing system on its potential disparate impact on communities of colour. We published the results of our work in *Significance*.[2] Quite a few people thought that this was a pretty salient piece of writing, and it led to me pursuing a career using statistical methods to evaluate algorithmic systems that are being deployed around the world.

This was an interesting shift, and it was not initially what I had envisioned for my career. However, it has actually proved to be a nice intersection between the things that I am passionate about, namely technology, civil rights and human rights.

I now work at the artificial intelligence (AI) company DeepMind. My technical title is senior research scientist, and I am part of the ethics and society team. In this team, our job is to think about the broader societal impacts of the work that DeepMind does. My work involves a mix of original research, as well as the testing and evaluation of projects that are in the domain of machine learning.

A lot of my work centres around fairness and bias in machine learning systems. So, for example, I recently published a paper with Sylvia Chiappa, looking at causality as a way to evaluate and measure bias or unfairness in a given system (bit.ly/3sfbdbo). I have also written about decolonisation in AI, thinking about the ways in which the work that we do has an impact on the historical artefacts of society and what that means, and what we want to do going forward. So, I get to wear quite a few hats, and the work ranges from very grounded technical work to thinking big picture and thinking more broadly about society and the emergence of these technologies.

One thing my career has taught me so far is the value of collaboration across disciplines. I would not describe myself as a statistician; I am a computational social scientist. But working with Kristian Lum at HRDAG helped me better understand statistics far more than reading a book about it ever did, because we were working together and having a dialogue about the nuances of our project. I feel that whenever you have an opportunity to collaborate with people who are outside of your specific discipline, then that is where you really get the most out of your own training, and where you also get to grow, in terms of expanding into new areas or by integrating your thinking with new domains.

In my job now, I am speaking with philosophers, computer scientists, engineers and statisticians, and so when I am writing a paper or doing an analysis, I feel that the work becomes richer because these collaborations lend new perspectives that I had not thought about. The nature of the problems that we are seeking to understand now is so complex, I think it is almost a necessity for researchers to collaborate and interact and engage with people from other disciplines and disciplinary backgrounds. ■

References

1. Lum, K. (2020) Career story. *Significance*, **17**(6), 40.
2. Lum, K. and Isaac, W. (2016) To predict and serve? *Significance*, **13**(5), 14–19.

ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

Gain new skills & advance your career

Are you looking to develop your skills?

Why not attend one of our training courses?

Our programme has both virtual and face-to-face courses, all of them designed to give you as much practical experience as possible.

Bayesian courses available:

- Introduction to Bayesian Statistics
- Introduction to Bayesian Analysis using Stan
- Bayesian Meta-analysis

Explore all our advanced statistical courses online at **rss.org.uk/public-courses**
or **Contact us on:** +44 (0)207 638 8998 training@rss.org.uk

Discounts are available for early bookings, group bookings and RSS members.
Become an RSS member and enjoy our full range of benefits.

rss.org.uk/join

Covid-19: a view from the sidelines

Katherine Hoffman is a biostatistician in the pulmonary and critical care team of a New York City hospital, who found herself part of the Covid-19 response when the outbreak first hit in March 2020. This is her story

When I received my master's degree in biostatistics from the University of Michigan almost 3 years ago, I possessed various skills. I could plot high-dimensional data and code complex statistical models, but more importantly, I knew how to consider many angles of a problem, choose the best approach, methodically analyse data, and present the final evidence as clearly and objectively as possible.

As I excitedly made plans to leave my small Michigan hometown for biomedical research in New York City, I had no inkling that there would soon be a time when I would need skills for which I was not explicitly trained. I did not realise I might one day need to provide immediate, incomplete answers using rapidly accumulating information on a new and unknown disease; did not foresee I would be asked to do this as bodies piled up in mobile morgues near my apartment faster than I could download new data. Yet that is exactly what happened, less than 2 years after I graduated, when New York became one of the early epicentres of the Covid-19 pandemic in the spring of 2020. As a biostatistician for a team of pulmonary and critical care physicians at a Manhattan hospital, I was part of the first wave of scientists to work on emergency response efforts.

I found these months of my life – filled with so much sadness and so little sleep – difficult to articulate even once the number of cases decreased to a manageable level in the summer. Like trying to recount a nightmare moments after you have been shaken awake, my description of reality, if spoken aloud at all, sounded nonsensical, dramatised, even untrue. The words of a New York City pulmonary and critical care fellow, Dr Colleen Farrell, echoed in my mind: "I find myself instead longing for the horror and devastation of this crisis to be seen and acknowledged for what it really is. I don't want to be soothed so much as believed"(bit.ly/38qMhvx).

While I knew my experience paled in comparison with that of those on the frontline, I decided to write about it. I share it because I believe that if you were "on the sidelines" of Covid-19 response efforts, as I was, pieces of my story are likely to resonate. Even though your own experience will have variations, you will understand the chronology of escalation, desperation, and exhaustion. If you were not involved, I hope you might live it, feel it, if

Katherine Hoffman is a biostatistician at Weill Cornell Medicine in New York City. She collaborates with physicians in the Pulmonary and Critical Care Division of NewYork-Presbyterian Weill Cornell Medical Center on all stages of their research: grant proposal, study design, statistical modelling, and manuscript writing.

only for a few minutes and through the eyes of one young statistician in her tiny Manhattan apartment.

There were and are so many of us: thousands of scientists crunching numbers and creating models and pipetting molecules to contribute to treatments and vaccines and decisions on resource allocation. We tried our best. Even when it was not enough – and it was, so often, not enough – we gave it our all, and our work carried its own emotional burden.

My story begins at the desk of my first job as a biostatistician at Weill Cornell Medicine. It is over a year before Covid-19 reshaped life as we knew it.

15 December 2018. My coworker is moving to California. She's a statistician for a group of pulmonary and critical care physicians at our New York City hospital, and I'm a statistician who's trying not to do too many things wrong, only 3 months into my first job out of school. "I think you'd be good with this research team," she tells me. "There's some really interesting studies on lung diseases." I nod, because that's what you do when you've been at your job for 3 months.

I take over her projects and start learning organ failure scoring systems, criteria for acute respiratory distress syndrome, and the differences between invasive and non-invasive mechanical ventilation. My close friend does cutting-edge cancer statistics, and I feel a bit resentful. Nobody ever wants to hear about the controversial definitions of sepsis at family parties.

As the months pass by, I slowly build my mental encyclopedia and begin to embrace my role as a pulmonary and critical care biostatistician. I do not consider – indeed, I could not have comprehended – how valuable this domain knowledge would soon become.

5 March 2020. A full year and 3 months later, I wake up very sick. It is the kind of sick where you cannot do anything but curl up on your bathroom floor and let *being sick* consume you. Too sick to read, too sick to sleep. I spike a fever and can hardly move for 2 days before I hobble to the doctor's office and nearly faint mid-exam. The doctor insists I stay until I drink an entire bottle of water. "Is there anyone to check on you at home?" he asks, concerned. No, no, I'll be fine.

By the end of the week my fever breaks and I'm back to work. It's early March, so "coronavirus?!" is everyone's first question. They're all joking, except the pulmonologists I work with. Nothing respiratory, I assure them. One isn't convinced. "Some young people don't feel short of breath. It *is* possible to have Covid-19 with no respiratory symptoms at all," she tells me. Months later, I'll read that as the headline of various news articles, but at the time, no testing is available to me.

17 March 2020. Barely 2 weeks pass before the number of confirmed Covid-19 cases explodes in New York City. Restaurants are instructed to close the day before St Patrick's Day, my birthday. I can't meet up with my friends anymore, so I cook macaroni and cheese and run to Central Park to watch the sun set behind skyscrapers. My grandparents call me, and they make *Happy Birthday* sound like a hymn from a Catholic mass and I laugh, and it is the only part of my day that feels like every other birthday.

While I'm leaving the park, my mum texts me that she hopes I had a good day. Any other year it would be strange for her to nearly miss my birthday, but this year she is working long hours. She's a nursing director back in Michigan and her hospital is already preparing for their own impending Covid-19 outbreak. The preparations will not be in vain.

As I jog home, I pass a sign asking former health-care workers to volunteer to take care of NYC Covid-19 patients. Before I began my career in biostatistics, I worked at a hospital caring for acutely ill patients, so I sign up without hesitation. My misguided logic is that the rising numbers of Covid-19 cases will make my critical care collaborators too busy to pursue their research, and this seems like the best way for me to help as the world descends into chaos. While I fill out the online contact form, I wonder what it will feel like to take care of patients again. I look up YouTube videos to refresh myself on drawing blood and inserting IVs.

How absolutely crazy that I thought my biostatistics training wouldn't be useful.

22 March 2020. I'm a pulmonary and critical care team's statistician, so *naturally* I am one of the first analysts at my hospital pulled into Covid-19 work. It starts with a text on a Sunday – the first of many – from a pulmonologist: "Where's the description of our ICU [intensive care unit] database, Kat?" Our informatics team is using the structure of the ICU database I work with as part of a Covid-19 tracking repository for our entire hospital. Within days, I am told to drop all of my other research projects for Covid-19 work.

The first request for me is straightforward: summarise the laboratory results from our first 300 Covid-19 patients. Three hundred patients at our hospital! That's insane, I think to myself. It seems only a week ago the news reports said there were 300 people in the entire city with Covid-19. I begin working through issues linking the databases, identifying missing information, and explaining critical care jargon to other analysts. Each morning I pull new data and watch the files grow exponentially larger.

There are countless questions flooding in from all over the hospital. Most of them revolve around "who will get intubated, and when?" My hospital, like so many other hospitals in NYC, is on track to run out of ventilators soon. My attendance becomes mandatory at multiple "risk prediction" meetings each week. I find myself in charge of extensive data cleaning and then writing code for models to answer vague and terrifying questions: we need to figure out which patients will "crash," who can be transferred, and, if we run out of ventilators, who has the best chance to survive.

I am a junior researcher, previously unconcerned with hospital operations, suddenly confronted with the task of providing rapid answers for potentially immediate decision making. I accept my new role with the utmost seriousness. My days, normally spent coding with double monitors at a proper desk, suddenly fill with online meetings from 8:30 a.m. until 5:00 p.m. from a laptop at my kitchen table. Each night after the meetings end, I take advantage of the relative quiet to code into the early hours of the morning.

For several weeks I use the long, uninterrupted hours of weekends to work, waking up with the sun and continuing on until at least 11 p.m., with few breaks in between. On some nights I send my mum "good morning" texts at 5 a.m. "Are you waking up early or have you not slept yet?" is always her first question. The next is, "No fever? No cough?" She is worried about me, living in the international epicenter of this pandemic, but I'm just as worried about her, working at a hospital every day. She informs me that my dad is sleeping in my old bedroom in case she brings the virus home.

Hospitals around the city begin to call me, wanting to know if I can still help care for Covid-19 patients, as I had previously volunteered to. I tell them: "I want to, but I can't, I'm so sorry, I'm helping with Covid data now." It sounds and feels inconsequential.

4 April 2020. My best friend and her sister are also nurses in Michigan. I videocall her to check in. She and her sister's units have become "hot floors": every room is filled with a Covid-19 patient. They were living with their parents, another sister, uncle, and cousins, but both have moved to rental accommodation for the foreseeable future. "It's so crazy here, Kitty," she tells me in a defeated voice. At the time, Michigan's case trajectory is second only to New York's.

One of her nursing friends has been hospitalised with Covid-19 and is on 6 litres of oxygen. I can't help but think about the prediction models I've been working on. I mentally run his characteristics through them. I know what my models would estimate his probability of intubation to be.

I listen to her talk about the N-95 masks they've been given. "Remember how they used to say those were one-time use?" she asks me. I do. "They started telling us they were good for the whole day, and then they said they'd be good for the whole week, and now they're saying we might have to start sharing." I wonder what data analyst, perhaps just like me, is crunching those numbers and feeding the information to hospital administration. "The virus is so terrible. I've never done so much post-mortem care, zipped so many body bags..." Her voice drifts off.

I feel guilty, on the sidelines. I see the raccoon eyes – the only part of their faces visible between hair caps and procedure masks – of the physicians I spend all day hopping on and off meetings with, and I desperately want to help. I cannot hold the hand of a Covid-19 patient, but I have all their data at my keystrokes: lab results, vital signs, and procedure codes. I see their inflammatory cytokines spike, I watch their oxygen levels plummet, I can tell you which organs are failing, who's on which experimental drug, and who's just been made Do Not Intubate and Do Not Resuscitate. I follow in horror, almost

in real time, the time-stamps of admission, intubation, death. I cannot compare this experience to physically caring for Covid-19 patients, but I feel haunted by it all the same.

I hole up in my tiny studio apartment in Manhattan for days at a time, listening to the wails of ambulances and pings of messages from my computer. I see only one friend with any frequency; we both live alone, 18 blocks from each other. She texts me often, asking to meet in Central Park. She suspects I am not doing well, and she is right. I walk with her all over the Upper East Side a few times a week, each of us donned in our black cotton masks. We try not to talk about Covid-19, but it's hard to avoid when our walks take us past the pop-up ICU tents and refrigerated trucks that stretch entire blocks – the overflow morgues for NYC's dead.

We try to time our walks so that we're outside at 7 p.m., when the city unites to cheer for health-care workers. If I'm not out walking with her, I climb religiously onto my fire escape to clap. Sometimes a man in the apartment across the street sings Sinatra: "I want to wake up in the city that never sleeps... New York, New York!" I've only lived here 2 years, but I miss "the city that never sleeps" so badly that it hurts.

Life continues in this way for me, with no real sense of time or distinguishing events, from mid-March until early May.

10 May 2020. It is Mother's Day, and my 50th straight day of working with Covid-19 data. At 11 p.m., my cell phone goes off. It is an ominous vibration against my kitchen table, where I am perpetually sitting with my laptop whirring. "Hi, honey... I just wanted to let you know that, mmm..." It's my mum, and her voice is cracking. I finish the sentence for her. "Aunt Peggy died?" I ask, sadly. "Yes." "Okay. Thanks for letting me know." I stare into the white brick wall in front of my kitchen table for so long that I start seeing multicoloured spots.

My grandfather's eldest sister, my Aunt Peggy, had begun showing telltale symptoms of Covid-19 and tested positive only a few days previously. She'd been without any visitors in her assisted living home for months due to isolation restrictions. She was royalty in our family; the red-lipsticked, always fashionably late, prized guest at every family party. She had an unforgettable, incredibly sweet voice, and I can still hear her words to me last Christmas. "How's *New York*, Katherine? I'm so proud of you." She was the first nurse in my family, and she influenced my mum to become a nurse, who influenced me to pursue medical research. The matriarch of our family left us on Mother's Day.

I spend the night trying to find a rental car company that will allow me to drive one-way from New York to Michigan. It can't be done; I am several weeks too late in my exodus from the city. I book a flight instead and leave a few days later on a near-empty plane to spend time with my family. I plan to stay in Michigan for 2 weeks, but I don't leave for 2 months.

20 September 2020. The leaves I watched bud in Central Park during my walks this spring are changing to red and gold. As I write this, I think of countless other ways I could attempt to explain what my tiny corner of the world was like during NYC's outbreak. Most are too personal to ever record. At the same time, it is difficult to share even the memories I have, partially because I know they are incomparable to those of the frontline workers who risked their lives every day.

My experiences of living and working in Manhattan during March, April, and May will stick with me forever. I hope there comes a day that I can meet in real life – mask-free – all the analysts, hospital administrators, physicians, residents, fellows, medical students, and data engineers I conversed with so frequently during the height of the outbreak. At the same time, I hope we never have to work together again. It is a wish that I fear will not come true.

Just this past week I attended a meeting with our informatics team. "It's good to 'see' everyone again," someone said. It's only half true; the circumstances that bring us to meetings together are never good. We discussed data structures for a possible second wave of Covid-19 in NYC as schools and indoor dining reopen. After the call, I felt an immense sadness, despite being in a much better place than when I left the city in May.

At the bottom of my heart, I don't know if I can handle another round of it all. Can you? ■

Disclosure statement
The author declares no conflicts of interest.

Strong public claims may not reflect researchers' private convictions

A survey indicates that some researchers are more modest about their findings than their published articles would suggest. **Johnny van Doorn**, **Eric-Jan Wagenmakers** and colleagues argue that authors should express this uncertainty openly

Chris Yang/Unsplash.com

How confident are researchers in their own claims? The nineteenth-century British mathematician Augustus De Morgan suggested that researchers may initially present their conclusions modestly, but afterwards use them as if they were a "moral certainty".[1] To prevent this from happening, De Morgan proposed that whenever researchers make a claim, they accompany it with a number that reflects their degree of confidence.[2] Current reporting procedures in academia, however, usually present claims without the authors' assessment of confidence.

Here we report the partial results from an anonymous questionnaire on the concept of evidence that we sent to 162 corresponding authors of research articles and letters published in *Nature Human Behaviour* (*NHB*). We opted for *NHB* because of its broad scope and because the majority of its articles include the main claim in the title (e.g., from the first issue, "Pathogen prevalence is associated with cultural changes in gender equality",[3] or "Attention modulates perception of visual space"[4]), which made it convenient to directly reference the claim in the questionnaire. We selected 129 articles with a claim in the title published between January 2017 and April 2020. The list of selected articles as well as a description of the selection procedure can be found in Appendix A of the online supplement (osf.io/zjnpm). We received 31 complete responses (response rate 19%). A complete overview of the questionnaire can be found in online Appendices B, C, and D.

As part of the questionnaire, we asked respondents two questions about the claim in the title of their *NHB* article: "In your opinion, how plausible was the claim *before* you saw the data?" and "In your opinion, how plausible was the claim *after* you saw the data?" Respondents answered by manipulating a sliding bar that ranged from zero (i.e., "you know the claim is false") to 100

Empirical disciplines do not ask authors to express the confidence in their claims, even though this could be relatively simple

(i.e., "you know the claim is true"), with an initial value of 50 (i.e., "you believe the claim is equally likely to be true or false").

Figure 1 shows the responses to both questions. The blue dots quantify the assessment of prior plausibility. The highest prior plausibility is 75 and the lowest is 20, indicating that (albeit with the benefit of hindsight) the respondents did not set out to study claims that they believed to be either outlandish or trivial. Compared to the heterogeneity in the topics covered, this range of prior plausibility is relatively narrow.

The lines in Figure 1 connect, for each respondent, their subjective assessment of prior and posterior plausibility; the positive slopes indicate that all 31 respondents believed that the data increased the plausibility of the claim from the title of their article (Wilcoxon signed-rank test, $W = 0$; $p < 0.001$; $BF_{-0} =$ 2,670,000; see osf.io/kd4ps). However, with a median of only 80, the posterior plausibility for their claims is surprisingly low. From the difference between prior and posterior odds we can derive the Bayes factor,[5,6] that is, the extent to which the data changed the researchers' convictions. The median of this informal Bayes factor is 3, corresponding to the interpretation that the data are three times more likely to have occurred under the hypothesis that the claim is true than under the hypothesis that the claim is false. A Bayes factor of 3 equals Jeffreys' threshold value for labelling the evidence "not worth more than a bare mention",[5] further underscoring the authors' modesty and/or seemingly weak convictions of their article's main claim.

The authors' modesty appears excessive. It is not reflected in the declarative title of their *NHB* articles, and it could not reasonably have been gleaned from the content of the articles themselves. Perhaps authors grossly overestimated the prior plausibility of their claims (due to hindsight bias); or perhaps they were afraid to come across as overconfident; or perhaps they felt that the title claim was overly general. It is also possible that authors were not sufficiently attuned to the response scale, although none of the respondents indicated that the scales were unclear.

Empirical disciplines do not ask authors to express the confidence in their claims, even though this could be relatively simple. For instance, journals could ask authors to estimate the prior/posterior plausibility, or the probability of a replication yielding a similar result (e.g., (non-)significance at the same alpha level and sample size), for each claim or hypothesis under consideration, and present the results on the first page of the article. When an author publishes a strong claim in a top-tier journal such as *NHB*, one may expect this author to be relatively confident. While the current academic landscape does not allow authors to express their uncertainty publicly, our results suggest that they may well be aware of it. Encouraging authors to express this uncertainty openly may lead to more honest and nuanced scientific communication.[7] ■

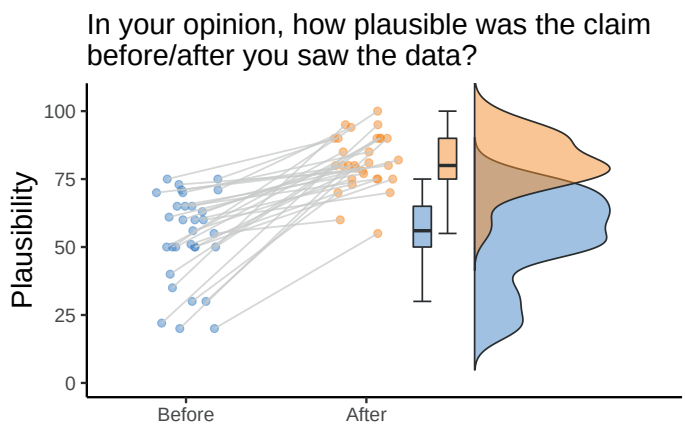


Figure 1: All 31 respondents indicated that the data made the claim in the title of their *NHB* article more likely than it was before. However, the size of the increase is modest. Before seeing the data, the plausibility centres around 50 (median = 56); after seeing the data, the plausibility centres around 75 (median = 80). The grey lines connect the responses for each respondent.

About the authors

Johnny van Doorn is an assistant professor at the University of Amsterdam. Don van den Bergh is a PhD student at the University of Amsterdam. Fabian Dablander is a PhD student at the University of Amsterdam. Noah van Dongen is a postdoctoral researcher at the University of Amsterdam. Koen Derks is a PhD student at Nyenrode Business University. Nathan Evans is ARC DECRA research fellow at the University of Queensland. Quentin Gronau is a postdoctoral researcher at the University of Amsterdam. Julia Haaf is an assistant professor at the University of Amsterdam. Yoshihiko Kunisato is an associate professor at Senshu University, Japan. Alexander Ly is a postdoctoral researcher at the University of Amsterdam and the Centrum voor Wiskunde & Informatica. Maarten Marsman is assistant professor at the University of Amsterdam. Alexandra Sarafoglou is a PhD student at the University of Amsterdam. Angelika Stefan is a PhD student at the University of Amsterdam. Eric-Jan Wagenmakers is a professor at the University of Amsterdam.

Disclosure statement

The authors declare no conflicts of interest.

References

1. De Morgan, A. (2003) *Formal Logic: The Calculus of Inference, Necessary and Probable*. Honolulu, HI: University Press of the Pacific. First published in 1847.

2. Goodman, S. N. (2018) How sure are you of your result? Put a number on it. *Nature*, **564**, 7.

3. Varnum, M. E. and Grossmann, I. (2016) Pathogen prevalence is associated with cultural changes in gender equality. *Nature Human Behaviour*, **1**, 0003.

4. Zhou, L., Deng, C., Ooi, T. L. and He, Z. J. (2016) Attention modulates perception of visual space. *Nature Human Behaviour*, **1**, 0004.

5. Jeffreys, H. (1961) *Theory of Probability* (3rd edn). Oxford: Oxford University Press.

6. Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

7. Editorial (2020) Tell it like it is. *Nature Human Behaviour*, **4**, 1.

Bias and meta-analysis: an exchange

In June 2020, *Significance* published an article by Geoffrey Kabat, titled "The two faces of meta-analysis". We received the following (edited) response from Professor Lianne Sheppard of the University of Washington, a co-author of a study cited by Kabat in his article:

> Geoffrey Kabat's article displays deplorable bias in its attempt to discredit our meta-analysis of observational studies of the effect of the herbicide glyphosate on non-Hodgkin lymphoma (NHL).
>
> In summarising our work, funded by the National Institutes of Health, Kabat's article trivialises its quality, fails to acknowledge the transparency of our reporting, and, most important, neglects to mention our biologically based *a priori* hypothesis that aimed to more clearly elucidate whether glyphosate may increase the risk of NHL. We asked whether there was increased risk of NHL among the most highly exposed workers in each study and reported a meta-relative risk of 1.41 (95% confidence interval (1.13, 1.75)) relative to unexposed workers.[1] We focused on this hypothesis because NHL development suggests that workers with higher exposures, including longer exposure duration, higher intensity, and longer latency, will show increased cancer risk, if indeed glyphosate causes NHL.
>
> The only attention to our scholarship in Kabat's article was to contrast it with the US Environmental Protection Agency's most recent meta-analysis findings (bit.ly/2XLKPQJ), which addressed a different scientific question, namely whether there is an increased risk among ever-exposed workers, a group which would include many whose exposure would be too small to induce risk no matter what the true effect of glyphosate. It is misleading for Kabat's article to present an apparent "apples to apples" comparison of two meta-analyses when these studies were asking completely different scientific questions.
>
> Furthermore, while implying our work was dominated by "subjective biases", Kabat's article fails to acknowledge his own potential biases.

Sheppard follows this up by pointing to a March 2019 report published by a website called US Right to Know (USRTK), which refers to Kabat's involvement with two groups: the Science Literacy Project (SLP) – the parent group of the Genetic Literacy Project (GLP) – and the American Council on Science and Health (ACSH). USRTK alleges that these groups "partner with the chemical industry on PR and lobbying campaigns", and this claim appears to be based on USRTK's reading of internal documents from the company Monsanto that include mention of both GLP and ACSH. These documents were released as part of litigation concerning the weedkiller Roundup, which was originally manufactured by Monsanto and whose main ingredient is glyphosate. One cited document, dated February 2015, mentions the "Genetic Literacy Project" as one of a number of "industry partners" that Monsanto planned to "engage" as part of its PR efforts to "[p]rotect the reputation ... of Roundup by communicating the safety of glyphosate" (bit.ly/3qzIu6y). Another cited document, also dated February 2015, contains the text of an email exchange between ACSH and Monsanto personnel in which glyphosate is discussed, as is a request "to secure Monsanto's renewed and continued funding of ACSH's ongoing work to provide a rational voice of scientific reason to the public" (bit.ly/2Ng3pgp).

On its website (bit.ly/39uPHhi), the GLP refers to allegations made by USRTK and states: "the GLP or its employees has not been offered or received any donation from Monsanto (or any of its employees) during the history of the nonprofit". GLP says that it is "funded by grants from independent foundations and charities" and that it "accepts tax-deductible donations from individuals and associations, but not from corporations", adding: "We have no affiliation ... with any corporation."

The ACSH does not report specific sources of funding on its website (acsh.org/financials) but gives a breakdown of funding by category for the most recent year: roughly 61% from individuals, 26% from private foundations, 9% from governments, 3% from trade associations, and 2% from corporations. It describes itself as a "pro-science consumer advocacy organization", adding that: "We are not a trade association. We do not represent any industry."

(USRTK reports its major donors at usrtk.org/donors – among them is the Organic Consumers Association, which has been its top donor in five of the last seven years.)

The GLP is part of the SLP, and the SLP was registered as a non-profit in 2015, according to the GLP website. Financial returns filed with the Internal Revenue Service for the years 2016, 2018 and 2019 list Kabat as an unpaid director of SLP. Kabat has since told *Significance* that he resigned as a director on 1 December 2020.

The ACSH website lists Kabat as a member of its board of scientific advisors.

We invited Kabat to comment on Sheppard's letter, and his affiliations with GLP/SLP and ACSH. He sent the following (edited) response:

> Professor Sheppard takes issue with my citing the Zhang *et al.* study[1] as an example of how the selection of risk estimates in a meta-analysis can affect the results. I stand by the example used. I cited this study because it combined the results of a high-quality, 20-year prospective study of 54,000 pesticide applicators (Agricultural Health Study [AHS]) with the results of five case–control studies, which are far inferior in quality. The Cochrane handbook on meta-analysis cautions against combining heterogeneous studies in a meta-analysis.[2] Zhang *et al.* then selected one of five risk estimates from the AHS, which yielded a 41% increase in risk of NHL in the highest exposure group. This was justified based on their *a priori* hypothesis that the highest-exposure group would be expected to have the highest risk. But the findings of the AHS do not support their *a priori* hypothesis. Neither the highest-exposure group nor any of the four exposure levels show any hint of

We welcome comments from readers – please email **significance@rss.org.uk**. Contributions should be no more than 300 words in length and clearly marked "for publication". Published responses may be edited to fit.

a positive association in any of five analyses, as pointed out by the US Environmental Protection Agency and by us.[3]

Sheppard refers to my "bias(es)" in the first and last sentences of her letter. However, there is a difference between bias and examining the range of different risk estimates from each study and judging which are the most credible. In our recently published paper,[3] we examine the effects of considering the full range of risk estimates in the studies on glyphosate and NHL.

Sheppard apparently refers to USRTK's smears to counter my arguments about substance, but there is no evidence to substantiate the allegations. The two organisations I've been associated with strive to provide high-quality reporting on important scientific issues, while eschewing ideology and demagogy.

References

1. Zhang, L., Rana, I., Shaffer, R. M., Taioli, E. and Sheppard, L. (2019) Exposure to glyphosate-based herbicides and risk for non-Hodgkin lymphoma: A meta-analysis and supporting evidence. *Mutation Research*, **781**, 186–206.

2. Deeks J. J., Higgins, J. P. T. and Altman, D. G. (eds) (2020) Chapter 10: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page and V. A. Welch (eds), *Cochrane Handbook for Systematic Reviews of Interventions*, version 6.1 (updated September 2020). Cochrane. bit.ly/3qyOlZK

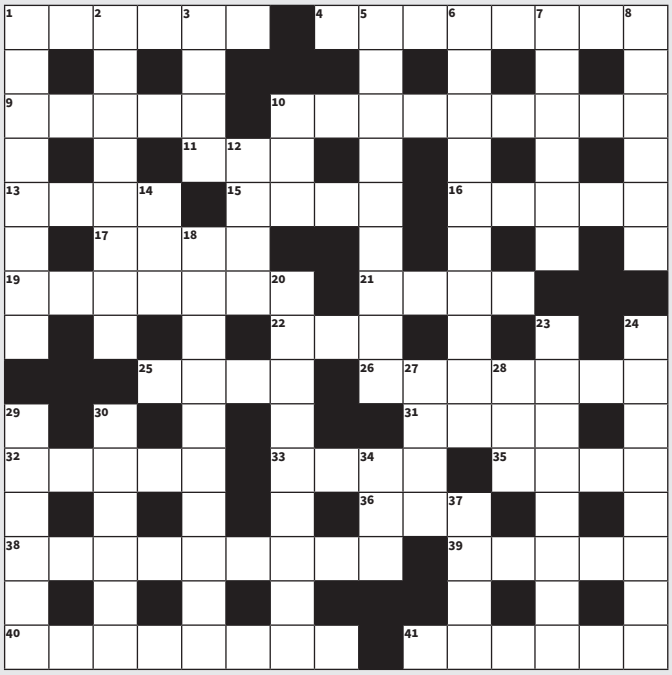
3. Kabat, G. C., Price, W. J. and Tarone, R. E. (2021) On recent meta-analyses of exposure to glyphosate and risk of non-Hodgkin's lymphoma in humans. *Cancer Causes & Control*. doi: 10.1007/s10552-020-01387-w.

The eyes have it

The last sentence in Joseph J. Locascio's letter, "A graph is worth a million words" (*Significance*, October 2020, page 47) reminded me of a similar remark included in the answer of one of my first-year undergraduate students, 1993–94, to a question on simple linear regression. It went: "Although regression can be done without ever looking at a scatter plot, that is the statistical equivalent of flying blind." I still quote it at opportune moments.

Michael Stuart
Trinity College Dublin

Wiley Prize Crossword: *Nobody Was Hurt* by Sam Buttrey



Send your solution to: *Significance* Crossword Competition, Royal Statistical Society, 12 Errol Street, London, EC1Y 8LX (photocopies are acceptable), or scan it and email to significance@rss.org.uk. The competition is sponsored by Wiley (wiley.com/statistics), who will give the winner £100 or $150 to spend on Wiley books. Closing date: 11 March 2021. The winner will be chosen randomly from the correct entries, and the solution published in a future issue.

Nobody got hurt, but something is happening at seven of the intersections in this grid. Entries at those spots will need to be placed carefully.

Across

1 Fodder for vegetarian Swedish group locked in cells after one escapes
4 Plated, could be salvaged in British style
9 Gets close to elements of genuine arson (5)
10 Particulars from outline provided regular pieces (9)
11 Article written with editor, in part (3)
13 Sage and hip after opening of symposium
15 Perhaps darn, or otherwise alter the hose tops (4)
16 After midnight passes, parent becomes something else (5)
17 I am Big Foot! (4)
19 Transport alien to paranormal (7)
21 Back in the day, you found an article at end of summertime (4)
22 Cold that is initially caught inside (3)
25 Be sure to decay into darkness
26 Mean potion swilled with bit of tequila (5, 2)
31 Small tail removed after introduction of surgeon (4)
32 Sounds like teams served yellow balls? (5)
33 I'll be chasing gallon, or a quarter pint (4)
35 Rascal in abuse of Indian sailor
36 Apple activator's central elements of curiosity (3)
38 Break down a pound (£1), used up (9)
39 Way to search for crass person (5)
40 Stuffing maker and artist mixed goulash
41 It's difficult when cobras slither onto you and me

Down

1 Kisses snatched? That's bad energy, creates concerns
2 Weird auras about west wind in monkey puzzles
3 Leader of uprising force sent away to Asia, perhaps (4)
5 Yes album covers excited one of the Pharaohs (9)
6 Rice in cognac? Disgusting, and might make you sick
7 Small insult (6)
8 Rat on wasteland (6)
10 Red, for one, found in Picasso's earliest (3)
12 Traveler's stove finishing Sterno (4)
14 Ink bottom of contract before asset emptied out (3)
18 Tiny movement mimics roses, but not very well (10)
20 People like Daniel Ortega can rig sauna to explode
23 Maybe button coat that falls apart (6, 2)
24 Search American hills for hot spaces
27 European capital, consisting of some pesos, lost (4)
28 Lawyer's not ensnared in civil unrest from the Right (3)
29 Course heading for beginning sailor, all busy, confused
30 Award for dental issue (6)
34 Where a golf ball is found, or an alternative story? (3)
37 Confused Navy's securing landing, initially, in a forest

Solution to December issue's crossword:
***Both Directions* by Sam Buttrey**

Theme: four words need to be "translated" from British to American English, and four from American to British English, as evidenced by the unclued entry CROSS ATLANTIC.

Across: **5** anag MAIL ELF; **7** HO THE AD; **10** C(ROW)D; **11** anag A TIE SCORE; **12** anag TIMES UP; **14** E-STATE; **20** EVE + rev DEN; **21** EV + anag STEER; **24** RES(ET)TING; **25** LO + RR + Y (enter TRUCK); **27** L(EVER)ET; **28** SURGE ON.

Down: **1** C(OO)K + IE (enter BISCUIT); **2** B(AND)IT; **3** anag ASSETS minus T + EVEN; **4** anag REVIVED + IT + A; **6** J(OH)N (enter LOO); **7** BON + NET (enter HOOD); **8** T + REASON; **9** pun NAP-PY (enter DIAPER); **13** anag VIP REVERES; **15** anag LARGE LIST; **17** J + UMP + E + R (enter SWEATER); **18** anag LEGION + (w)AS (enter PETROL); **19** ST(OK)ING; **22** E + N + TIRE; **23** anag TO REVEAL (enter LIFT); **26** US + E.

Winner: J. M. Bell, via email

The secret statistician

Statistics for pleasure, if not profit

Renée Fisher/Unsplash.com

About 20 years ago, I was diagnosed with type II diabetes. This was not a great surprise as it runs in my family, but it meant a lot of needles and a lot of pills and, two decades later, it meant me having to hide away from infection with Covid-19, which really has it in for elderly diabetics of type II.

Every morning since my diagnosis, I have pricked my finger to squeeze out a drop of blood which a little electronic device uses to measure my blood glucose concentration. To a statistician a daily measurement means a regular source of data, and I have recorded these fasting glucose levels on my laptop. I draw graphs and have done some statistical analyses, but my diabetes changes over the years, as does the medication that goes with it. I now have such a long sequence of something that varies in a way that would be difficult to model. Analysis would be difficult.

It occurred to me, though, that one analysis which would be fairly straightforward would be to answer a question my wife posed: does curry increase my fasting glucose the following day? I started recording some items of diet in 2018 and so had for most days a log of whether or not I had eaten curry. I thought that I could ignore the effects of progressing disease and of changing medication if I compared the glucose measured on the morning after curry with the mornings before and after that. I did this and found that the mean difference, curry minus no curry, was 0.24 mmol/l, 95% confidence interval 0.07 to 0.44, $p = 0.008$. To put that in context, over the three years my fasting glucose had mean = 10.2, SD = 1.9 mmol/l (which is far too high and why I have recently started on insulin). So, the difference is small, but real. This could be for many reasons, but it showed that my wife was right (and that pleased her). Maybe it was the biochemical effect of one of the spices, maybe the effect of the carbohydrate in poppadoms (I don't eat rice). There are many possibilities and many factors to consider. But it does illustrate the fascination of statistics!

Collecting data on oneself is an interesting activity for a medical statistician. I have learned a lot about my condition, and I recommend it to any of you in a similar position. At the very least, it has given me something else to do during the weeks and months of Covid-imposed lockdown.

Which brings me to the real point of this column. This is my swan song, my last as the secret statistician (and "Dr Fisher" before it). I have held this berth since 2009 and I retired from my professional post five years ago, so I am getting a bit out of touch, especially being in isolation for most of the past year. I have really enjoyed having this regular soapbox, and I hope to be gracing the pages of *Significance* under my own name from time to time. But I am stepping aside to allow a new, possibly younger, curmudgeon a chance to address you.

Farewell. Stay safe. And maybe go easy on the curry. ■

Editor's note

I would like to thank the secret statistician for his years of contributions to *Significance*. Reading his columns back before I became editor of the magazine, I found them to be an insightful and witty introduction to the mindset of a statistician. And they clued me in to the fact that those trained in statistics tend to see the world very differently from those who are not so trained. So, sad as we are to say goodbye and good luck to the secret statistician, we do look forward to introducing readers to new voices, and new perspectives, in upcoming issues.

Significance is published in 6 issues per year. Institutional subscription prices for 2021 are: Print & Online: US$527, €326 (Europe), £259 (UK), $562 (Rest of World). Prices are exclusive of tax. Asia-Pacific GST, Canadian GST/HST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to onlinelibrary.wiley.com/library-info/products/price-lists/payment. The price includes online access to the current and all online backfiles to 1 January 2004, where available. For other pricing options, including access information and terms and conditions, please visit onlinelibrary.wiley.com/library-info/products/price-lists. Terms of use can be found at onlinelibrary.wiley.com/terms-and-conditions.

Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered at Place (DAP); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers Free of Board (FOB) our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability.

Single issues from current and recent volumes are available at the current single-issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 351 Fairview Avenue – Ste 300, Hudson, NY 12534, USA. Tel: +1 518 822-9300, Fax: +1 518 822-9305, Email: psc@periodicals.com.

JOIN US TODAY!

EXPAND
your professional network

EXPLORE
leadership opportunities

STAY CURRENT
in your area of expertise

Learn more at *www.amstat.org/join.*

ASA AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics®

ROYAL STATISTICAL SOCIETY

DATA | EVIDENCE | DECISIONS

RSS International Conference 2021

FOR ALL STATISTICIANS AND DATA SCIENTISTS. **ALL WELCOME**

MANCHESTER

6-9 September 2021

Submissions open for talks and posters

The RSS Conference offers a broad and varied programme of talks and workshops not found at any other UK statistical conference.

Now is your chance to contribute to that programme

If you've been involved in projects, new developments or research, why not share your work at this prestigious conference?

You can contribute in three ways:

- 20 minute talks
- 5 minute rapid fire presentations
- Posters

Deadline for talk submissions: **6 April 2021**

CONNECT • SHARE • LEARN

For more details visit:
rss.org.uk/conference2021
#RSS2021Conf @RSSAnnualConf