

Westward Ho! Musing on Mathematics and Mechanics

Alan Champneys CMath FIMA, University of Bristol

The so-called Bristol heart scandal in the mid 1990s was a watershed moment for the collection and analysis of medical statistics in the UK. In his somewhat sobering latest piece in the Westward Ho! series, Alan Champneys considers the implications 20 years on from the resulting public inquiry. On a par with Florence Nightingale's analysis of deaths during the Crimean war and Ronald Fisher's promulgation of the randomised controlled trial, it is argued that David Spiegelhalter's contribution to the Bristol public enquiry led to a revolution in drawing inference from routinely collected medical data. We also learn of a recent initiative in which mathematical scientists are working with clinicians to use the data to understand at a systemic level the consequences of COVID-19 on cardiac waiting lists.



The heart of the matter

My eldest son was born in the early summer of 1995, just around the time that a major medical story was erupting in the local press. The so-called Bristol heart scandal involved the high death rates of babies following cardiac surgery at the Bristol Royal Infirmary (BRI) from 1984 to 1995. Our son was born in the Royal United Hospital, Bath, but Bristol was the major care centre for any baby born with severe complications. Indeed, one of the babies born to parents in our National Child-birth Trust antenatal group, was born with a heart defect. Oscar (not his real name) was given a poor prognosis and needed emergency surgery at the Bristol hospital in question. Oscar is now a healthy adult and a father himself.

The Bristol heart scandal may not seem like a natural subject for an article in *Mathematics Today*, but unlike some of the other gruelling medical scandals to hit the national headlines, there would appear to be some subtlety to what happened. While naturally a sensitive subject, this is not simply the story of a rogue doctor (or doctors). Instead, at the heart of the matter is a question of data science. In an era when recently pioneered life-saving treatments for infants were being rolled out across the world, how was data being used to inform best practice? More specifically, in the case of Bristol, how was data being used to identify practice that was failing to meet the highest standards? And, crucially, how was this being used to inform management practice and patient choice within healthcare?

The culture change that arose as a result of the subsequent public inquiry can be argued to be a triumph of statistical analysis over an overconfident culture in which senior doctors' expertise and judgements were accepted without question. Not only have the Bristol Heart Institute and the Bristol Children's Hospital become some of the most reviewed medical units in the world, but detailed, patient-level outcomes are now collected from every cardiac hospital in the UK. The National Institute

for Cardiovascular Outcomes Research (NICOR) data set (www.nicor.org.uk) is made publicly available for the use of hospitals, care commissioners, researchers and, crucially for patients and their carers to enable all to make informed choices. But I am getting ahead of myself.

Of course, the collection of data and use of statistics has long lain at the heart of medicine, going back at least to the 1850s and the work of Florence Nightingale. The delightful article by Noel-Ann Bradshaw CMath FIMA in these pages [1] celebrated Nightingale's 200th anniversary. As Bradshaw argues, rather than as the founder of modern nursing, Nightingale should rightfully be remembered as the pioneer of medical statistics and the use of data visualisation to enable social reform. She more or less invented the infographic, using the radial-area plot (Figure 1) to illustrate how many more soldiers in the Crimean war were dying of preventable disease than of their injuries.

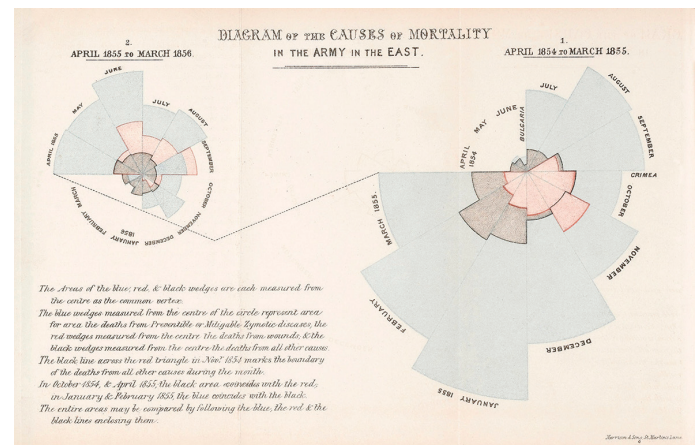


Figure 1: Florence Nightingale's radial-area plot depicting the cause of death by month of soldiers during the Crimean War. Credit: Wellcome Collection, CC BY 4.0.

Nightingale was a master communicator who throughout her long life used statistics and data science to improve society. The popular image of the lady with the lamp touring wards to visit sick soldiers (Figure 2) was, seemingly, a carefully managed public relations strategy that enabled her message to be heard. Perhaps her most enduring achievement was her use of data in lobbying the government to legislate for improvements in public sanitation. It has been claimed that these innovations alone led to an increase in life expectancy in the UK far beyond any medical intervention until the discovery of antibiotics.

Another pivotal use of mathematics in medicine has been the widespread adoption of the randomised control trial. Although dating back to the work of Scottish doctor James Lind in 1751, the importance of randomisation owes much to the statistician

and geneticist Ronald Fisher working at Rothamsted Experimental Station in the 1930s. The importance of such trials has recently been brought to the public's attention through the widely reported rates of effectiveness of the various vaccines developed to conquer COVID-19. But that is another story.



Figure 2: Florence Nightingale; her popular image as the Lady with the Lamp. Credit: Wellcome Collection, CC BY 4.0.

In modern medicine, the vast majority of data is not deliberately collected as part of a randomised control trial. Far, far more data from which we could be drawing inference is being routinely collected and thrown away – from the many ‘beep-beep machines’ in intensive care units, to the specifics of individual ‘patient pathways’ of treatment for complex chronic conditions. Modern clinicians are beginning to realise that there is value in this data, but the challenge for us mathematical and data scientists is to collaborate, just as Florence Nightingale did, to enable the data to be used to improve patient outcomes and quality of life. And this is where I believe the repercussions from Bristol in the 1990s have been pivotal.

Much has been written about the Bristol heart scandal, which I cannot do justice to here. The full report of the public enquiry, which ran from 1998 to 2001, makes sobering reading [2]. It is important, though, to understand the background.

During the 1980s and 1990s, the UK National Health Service (NHS) was in a state of flux. A key change came through the 1990 NHS and Community Care Act. The BRI became part of one of the first new hospital trusts, which operate like independent corporations and are led by a chief executive. The act also removed crown immunity from public sector healthcare providers. A new litigious, ‘ambulance chasing’ culture began in earnest. This exacerbated a culture that suppresses mistakes, which inevitably occur in any highly technical domain, and with them the potential for learning and improvement. See, for example, the book [3] on how both the medical and criminal justice professions could learn from the black boxes used in the aviation industry. Investigations aim to ensure that no two accidents should ever occur due to the same cause.

At the same time, cardiac surgery was undergoing a revolution. In the preceding two decades, new surgical techniques had been developing apace across the world, accelerated by the first successful adult heart transplant in 1967. Transplantation is just the tip of an iceberg, being a rarely performed eye-catching gross intervention. It, perhaps, masks a bigger revolution, that of open-heart surgery performed on babies born with major congenital

heart defects. Such defects, which develop in the womb and affect the structure and function of the heart, are estimated to occur in about 1% of live births.

Some more severe defects, such as dextro-transposition of the great arteries (d-TGA) – in which oxygen-depleted blood is sent around the body while oxygen-rich blood is pumped back to the lungs – could previously be treated only palliatively. This rare cause of blue baby syndrome (of which there are many other less serious causes) is estimated to affect about 1 in 3500 live births. Such babies were not expected to live long. The complex arterial switch procedure in which the two main arteries are severed, swapped over and reattached while the heart is bypassed, only moved from the realm of experimental surgery to become routine in the late 1980s.

Concerns into the quality of paediatric open-heart surgery at the BRI date back to the mid 1980s, when the hospital was first designated a supra-regional centre for such work. Serious concerns began to be raised in the 1990s by anaesthetist Dr Stephen Bolsin, who had been appointed in 1989 to work with the Bristol paediatric cardiac surgeons. He noticed how much longer babies were being kept on bypass during the more complex surgeries, compared with those he had witnessed during his previous career in London. Outcomes were worse too, both in terms of mortality (death) and morbidity (non-fatal negative outcomes). Also the Bristol team were having very little luck with the arterial switch procedure.

Things were exacerbated by staff shortages and the fact that the iconic Bristol Royal Hospital for Sick Children was separate from the main hospital building where heart surgery was performed, and was no longer fit for purpose. There were no dedicated infant heart surgeons who did not also operate on adults, and no specialist intensive care ward for the infants undergoing these procedures.

Unfortunately, the culture within the NHS at the time did not allow Dr Bolsin's concerns to be properly dealt with. First, there was no definitive source of data. There was lots of data, but it was held in individual databases, by different bodies with little or no cross-referencing. Also, there was, frankly, an ‘old boys club’ among surgical consultants in which there was no systematic requirement for performance evaluation nor skills development.

Over several years, Bolsin collected his own evidence and started to publish his findings. Management culture at the time was such that these findings were largely not acted upon. Part of the complication was that the trust's chief executive was risk averse and deferred all decisions to the medical team, but the medical director at the time was one of the very surgeons whose competency was being called into question.

Things came to a head in January 1995 through the case of poor Joshua Loveday. By this stage, it was clear that something was going wrong with the more complex surgeries at Bristol. A partial plan was already in place; one of the two surgeons involved stepped back from infant work, while it was decided to hold off on more complex surgeries until after a new dedicated paediatric heart surgeon took up their appointment later that year.

Joshua had been born in Gloucestershire, 16 months previously with d-TGA. As was common practice at the time, he had received simple palliative surgery at Bristol as a newborn, with a view that he would undergo an arterial switch within a year or so, but his condition was deteriorating. Surgery was scheduled for 12 January. Various doctors within the hospital advised

that the surgery should not go ahead. Unknown to Joshua's parents, an emergency meeting of cardiac specialists was convened on 11 January in which unchallenged 'back of an envelope' statistics were presented on the chances of success. It was decided by a majority that surgery should go ahead, with Dr Bolsin's being the only dissenting voice.

Sadly, the next day Joshua died in the operating theatre.

What followed next is well documented. An internal inquiry, based on a one-day visit by a specialist from another hospital, started to expose the ongoing problems that were staring everyone in the face. Everyone, that is, who was prepared to look at the data Dr Bolsin had collated. The report was leaked to the press and became a national story. A medical tribunal was called. Despite the supposed pause in complex open-heart surgery, there was another death, during a different procedure to correct an atrial septal defect (hole in the heart). What was then the longest medical tribunal in British history resulted in two doctors being struck off and a third having restrictions placed on their practice.

The full story emerged only after a full public inquiry chaired by Ian Kennedy QC, which finally reported in 2001 [2]. Until then, there was considerable uncertainty about whether the sanctioned doctors had been made scapegoats. Dr Bolsin also felt he had been victimised as a whistle-blower and, considering himself unemployable in the UK, had emigrated to Australia. Questions were asked as to whether he had acted unreasonably, selectively exposing inevitable failures in risky complex surgery, while failing to acknowledge other pioneering successes by the Bristol cardiac team.

The answer, of course, was statistics. The definitive piece of evidence presented to the inquiry was by a team of statisticians, led by David Spiegelhalter OBE FRS FIMA, then employed at the Medical Research Council Biostatistics Unit in Cambridge. The paper [4] that is based on their evidence is a textbook example of how to draw inference from messy and often conflicting sources of data, involving a complex range of outcomes over a range of procedures, during a period when considerable advances were being made in surgical techniques.

Their solution was to segment the data into comparable chunks. They divided the study period from January 1984 to April 1995 into three epochs, using an accepted governmental classification of 13 different surgery types, and separating post-natal surgery (under 1 year of age) from other child surgery (1–15 years old). Crucially, the imperfect sources of data for outcomes at Bristol were compared with the same sources at the 11 other English hospitals performing paediatric cardiac surgery.

The conclusions were stark. The authors of [4] are quoted in [2, p. 241]:

The single most compelling aspect of the data is the magnitude of the discrepancy between the outcomes observed at Bristol and those observed elsewhere. For children aged under one year undergoing open surgery between 1988 and 1994, the observed mortality rate at Bristol was roughly double that observed elsewhere in 5 out of 7 years. While the national trend over this period was for mortality rates to fall substantially, no such trend was seen in the Bristol results.

Particular emphasis was placed on the analysis of data from 1991

to 1995, which was the most reliable. They concluded that the increased mortality rate among children under one during that period was substantial and significant. In total, there were between 30 and 35 excess deaths at Bristol compared with what would have been expected at a typical comparator hospital. Powerful statistical methods were used to check that this conclusion was robust against adjustments for differing patient risk factors and other confounding variables.

David Spiegelhalter became the Winton Professor of the Public Understanding of Risk at Cambridge. Of note to this feature is that he is a West Country lad, being born and educated in the Devon town of Barnstaple. Rather like Florence Nightingale, he has established a reputation as both an advisor to official bodies on the use of data to quantify risk and for his work with the general public.

His book with Michael Blastland [5] attempts to give an entertaining view of what is a normal and acceptable level of risk. He has promoted simple concepts such as the *micromort* (a one in a million chance of death) and the *microlife* (a 30-minute reduction of life expectancy). For example, he did his first skydive as a 58-year-old, realising that at 7 micromorts, this was about equivalent to a 40-mile motor bike ride and a thousand times less than the chance of dying of natural causes before his next birthday. For some activities, such as cycling, he argues that the health rewards outweigh the risks. His most recent book [6] attempts to lift the lid more generally on how to use statistics to quantify risk and to make informed decisions. In 2014, David Spiegelhalter was knighted for his contribution to statistics.

The Bristol public inquiry [2] made many recommendations to avoid systemic failings within the NHS that led to poor performance at Bristol going unchallenged for so long. The key is data. Specifically, the drawing of comparisons and inference from routinely collected data. A new culture has emerged across the acute functions of the NHS of using such data to seek continual improvement. Within heart surgery, the correction of atrial septal defects, arterial switch and other open-heart surgeries on newborns have become almost routine with diagnosis often taking place in the womb. While still complex operations, published international statistics suggest success percentages in the high 90s.

Just as the inquiry report was being made public, a new purpose built Bristol Children's Hospital building was opened (Figure 3). Rather than up the hill and across the road, it is joined to the main building of the BRI where cardiac and other major surgery occurs. The old hospital building was made available to the University of Bristol, who used it as decant space for those affected by an engineering faculty building project in the Queen's Building. A chilling reminder was, therefore, given to many of us every day, as our PhD students worked in, and much of our teaching took place in, what had previously been triage and recovery wards for the Bristol heart babies.

While health outcomes have undoubtedly improved as a result of adopting the Bristol inquiry's recommendations, there remains at least one controversy. It was decided the inquiry would consider only 30-day mortality rather than morbidity more generally. It seems that data on the number of babies who suffered brain damage, went on to develop other severe conditions or died as an indirect consequence of the poor practice in the Bristol cardiac unit have never been recorded. It has been estimated that there may be more than 150 such cases. Lawyers acting for parents of those affected are still pursuing litigation.



Figure 3: The old Bristol Children's Hospital (left) and the new hospital, which opened in 2001.

The human stories can be particularly harrowing. Haunted by his son's death, Joshua Loveday's father's mental health suffered greatly. Having never previously been involved in crime, he participated as a lookout in an armed robbery. Racked with guilt, he gave himself up, received a three-year jail sentence and was found hanged in his cell a month or so later. He is thought to be one of four 'Bristol parents' who have committed suicide.

Monitoring the success rates of complex surgery is, in fact, just the tip of the iceberg in the data revolution that is now underway in healthcare. It has been estimated that 80% of the UK's health budget now goes on managing chronic conditions, rather than acute interventions. The model of health in which we go to a doctor and get a cure is no longer appropriate. Dementia, arthritis, COPD and even diseases like HIV and many cancers are now treated as conditions to be managed and lived with. But what is an optimal treatment strategy? Double blind control trials are still the gold standard, but they are expensive and often too blunt a tool for assessing marginal gains from minor adjustments to interventions. Could we instead draw inference from the swathes of data that are routinely collected across the healthcare sector?

And then there are real-time system-level issues – predicting case loads, planning hospital beds, understanding when a patient is 'ready' to be discharged, remotely monitoring at-risk patients, self-diagnosis and treatment – which could be informed by the real-time analysis of routinely collected data. In recent years, I have been on the fringes of a couple of interdisciplinary teams that are trying to use machine learning to understand such questions within intensive care wards [7, 8]. I get the impression, though, that there is much more to be done.

At the start of UK's first COVID-19 lockdown, in late March 2020, I became involved with an informal initiative called the Virtual Forum for Knowledge Exchange in the Mathematical Sciences (V-KEMS); see www.vkemsuk.org. This is a collaboration between the Newton Gateway, part of the Isaac Newton Institute in Cambridge, the International Centre for Mathematical Sciences based in Edinburgh and the UK Knowledge Transfer Network, together with Chris Budd OBE CMath FIMA from the University of Bath, Rebecca Hoyle FIMA from the the University of Southampton and independent consultant Jo Jordan FIMA. Together, we have run a whole series of virtual mathematical modelling study groups, in collaboration with a wide variety of public and private organisations, related to urgent national problems

arising from the interruption to the economy and society as a result of the pandemic. These workshops and related activities have involved many hundreds of PhD students and faculty from across all branches of the mathematical sciences in the UK and beyond, as well as domain experts and policymakers from many different fields.

In early February 2021, we held a study group on Modelling Solutions to the Impact of COVID-19 on Cardiovascular Waiting Lists [9]. Clinicians Chris Gale and Ramesh Nadarajah from the University of Leeds, Mamas Mamas from the University of Keele and Ben Gibbison from the University of Bristol have been instrumental in collecting data on the perturbation to cardiac waiting lists during the pandemic. The question we all sought to address is how we could draw inference from this data to minimise patient suffering. It seems that cardiac surgery is just the acute end of a much larger branch of medicine, cardiology, which looks into the management of risk among patients having progressive diseases of the cardiovascular system.

It is not quite as simple as saying that there is a backlog of cancelled surgery, which must be dealt with. Rather, there have been many missed appointments, which are missed opportunities for early intervention that could improve the prognosis for the estimated almost a million patients in the UK suffering from heart disease. The sub-group I became part of considered heart failure, a generic name for a range of conditions in which the heart fails to pump sufficiently well. This is a completely separate condition from heart attack, or myocardial infarction to use its proper name, which is the death of heart tissue caused by a clot in the arteries supplying the heart muscle itself. Although to be fair, many problems with the heart are interrelated.

The answers to our questions would seem to lie in operational research. Can we use the data collected both before and after the pandemic to understand how the healthcare system for treating cardiovascular disease operates? Crucially, can we find an optimal way of dealing with the backlog? Here, one has to be so careful in designing a suitable objective function. A natural choice would be to try to minimise micromorts. That is, to maximise so-called quality adjusted life years (QALYs). However, a natural conclusion from such a maximisation would be not to treat those who are most sick, as their life expectancy is probably least among all patients awaiting treatment. By not treating them, this would allow more resources to be spent on the not yet seriously

ill for whom early intervention might have the biggest impact.

Clearly, such an approach would be unethical. But, one could easily argue that treating the most sick first while the least sick get sicker is also unethical, because that would not clear the backlog and would be least optimal in terms of QALYs.

The preliminary report from the study group, like all V-KEMS reports, represents the work of many individuals and contains the important caveat [9, p. 3]:

This report contains preliminary findings that have not been peer reviewed. The findings are intended to provoke further study and policy discussion and should not be treated as definitive scientific advice in response to the COVID-19 pandemic.

Crucially, as the clinicians have pointed out to us, the important step is to develop practical implementable decision support tools that can lead to interventions that will improve patients' lives. And soon. To appreciate the nature of the challenge, interested readers might like to listen to the podcast organised by *Plus Magazine* [10].

As I write, Christine Currie and Alex Heib from the University of Southampton and Lucy Morgan and Alan Wise from the University of Lancaster are working with the clinicians at Leeds on a discrete event simulation tool based on the Leeds data that can, hopefully, be rolled out to other hospitals. Also Jess Enright from the University of Glasgow, Lars Schewe from the University of Edinburgh and Rebecca Hoyle have built models that seek to optimise cardiology patient flow more generally. They are seeking to recruit a researcher to work with the NHS improvement team in the West Midlands who have designed a systems dynamics model of patient waiting lists post-pandemic. Both groups really want to make a difference.

I cannot stress enough the human suffering that cardiovascular disease can bring to patients and their loved ones. Moreover, I certainly do not want to give any impression that mathematicians and data scientists are in any way smarter than medics. My professional interactions in the last few years with clinicians have shown me just how ingenious and solution-focused they are. Every day they have to make life or death decisions, based on a multi-disciplinary skill set, piecing together evidence and trusting their knowledge and technical skill. The Hippocratic oath seems

central to all the doctors I know, as they continually strive to do the best for their patients.

Nevertheless, I would like to dedicate this article to all those who have been affected in any way by this sensitive subject. At a personal level, in January of this year, my brother David died of undiagnosed heart failure. He passed away peacefully in his sleep, despite being a seemingly completely fit and healthy 55-year-old. Rest in peace, big brother!

REFERENCES

- 1 Bradshaw, N.A. (2020) Florence Nightingale: 200 years since her birth and we are still making the same errors with data, *Math. Today*, vol. 56, no. 2, pp. 65–68.
- 2 Kennedy, I. (2001) *Learning from Bristol: The Report of the Public Inquiry into Children's Heart Surgery at the Bristol Royal Infirmary 1984–1995*, The National Archive, tinyurl.com/Bristol-Inquiry-archived (archived 11 August 2009).
- 3 Syed, M. (2015) *Black Box Thinking: Marginal Gains and the Secrets of High Performance*, John Murray, London.
- 4 Spiegelhalter, D.J. et al. (2002) Commissioned analysis of surgical performance using routine data: lessons from the Bristol inquiry, *J. Roy. Statist. Soc. A, Part 2*, vol. 165, pp. 191–231.
- 5 Blastland, M. and Spiegelhalter, D.J. (2013) *The Norm Chronicles: Stories and Numbers about Risk*, Profile Books, London.
- 6 Spiegelhalter, D.J. (2019) *The Art of Statistics: How to Learn from Data*, Penguin, London.
- 7 McWilliams, C. et al. (2019) Towards a decision support tool for intensive care discharge: Machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK, *BMJ Open*, vol. 9, p. e025925.
- 8 Shillan, D. et al. (2019) Use of machine learning to analyze routinely collected intensive care unit data: A systematic review, *Crit. Care*, vol. 23, p. 284.
- 9 V-KEMS (2021) *Modelling Solutions to the Impact of COVID-19 on Cardiovascular Waiting Lists*, V-KEMS Study Group Report, tinyurl.com/VKEMSWaitingLists.
- 10 Plus Magazine (2021) *Reducing NHS waiting lists in times of COVID*, Podcast, plus.maths.org/content/reducing-nhs-waiting-lists-times-covid.