Professor Sylvia Richardson

Accepted: 31 August 2022
DOI: 10.1111/rssa.12957

ORIGINAL ARTICLE

Statistics in times of increasing uncertainty

Sylvia Richardson

Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK

Correspondence
Sylvia Richardson, Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, UK.
Email: sylvia.richardson@mrc-bsu.cam.ac.uk

Abstract
The statistical community mobilised vigorously from the start of the 2020 SARS-CoV-2 pandemic, following the RSS's long tradition of offering our expertise to help society tackle important issues that require evidence-based decisions. This address aims to capture the highlights of our collective engagement in the pandemic, and the difficulties faced in delivering statistical design and analysis at pace and in communicating to the wider public the many complex issues that arose. I argue that these challenges gave impetus to fruitful new directions in the merging of statistical principles with constraints of agility, responsiveness and societal responsibilities. The lessons learned from this will strengthen the long-term impact of the discipline and of the Society. The need to evaluate policies even in emergency, and to strive for statistical interoperability in future disease surveillance systems is highlighted. In my final remarks, I look towards the future landscape for statistics in the fast-moving world of data science and outline a strategy of visible and growing engagement of the RSS with the data science ecosystem, building on the central position of statistics.

KEYWORDS
Covid-19, data science, impact of statistics, policy evaluation, statistical agility, statistical expertise, statistical interoperability

[The address of the President to be read to the Royal Statistical Society on Monday, November 14th, 2022]

1 | INTRODUCTION

I was deeply honoured when I heard that I had been nominated as President Elect of the Royal Statistical Society (RSS) and remain humbled by the responsibility that has been bestowed on me. At the time I was appointed, in summer 2019, I had little idea of what an eventful period my presidency would be. I have been a fellow of the Society for over 40 years, since December 1979. I had just become a new lecturer in the statistics department of the University of Warwick and was enthused by the lively debates that were going round the department at the occasion of the upcoming RSS discussion papers. Listening to these meetings made me realise how stimulating and impactful for the discipline is the art form of robust but courteous statistical fencing expertly practised during such events. Organising open informed discussions and creating conditions for the melting pot of ideas and theoretical understanding with actions, education and practice has been one of the hallmarks of our Society though its illustrious history. As I got to know the Society's workings more intimately by being a member of the Research section (1995–1998), of Council (1998–2002) and by becoming Honorary Secretary for Publications (2003–2006) a period during which I took an active role in the creation of the magazine 'Significance', I gradually built my understanding of the breadth of the Society's vision of 'a world with data at the heart of understanding and decision-making' and a fuller appreciation of how this vision has been articulated and enabled throughout its long history, building on the rich diversity of interests and skills of all its members and friends of statistics.

During 2019, I thought about the massive expansion in 'volume, velocity and variety'[1] of data that we had experienced during the last decades, the transformative changes in the use and role of data in society and the ensuing challenges for statistics and the RSS. I decided that during my presidency, I would pursue and facilitate renewal and strengthening of the Society's engagement with the data science ecosystem and with data scientists. I had been struck by the analysis and warnings in the National Science Foundation (NSF) report 'Statistics at the crossroads' and thought it would be important to examine our activities and strategy with regard to some of the conclusions of the report and our collective experience. Little did I know at that time that at the beginning of 2020 a momentous societal tsunami would hit us in the shape of a pandemic created by a new fast spreading virus, the SARS-Cov-2 virus, which would make extraordinary threats and demands on the whole world, bring into focus the fragility and uncertain future of global health, and solicit a massive international scientific effort at an unprecedented pace. The pandemic changed the course of my presidency. Following the 24 March 2020 council meeting, the week of the first national lockdown, the RSS convened a Covid-19 Task Force, which I volunteered to co-chair jointly with David Spiegelhalter. From the beginning, questions and queries kept recurring on requisite data infrastructure, data access and data quality, on ways of implementing swift evaluations of treatment or policies, and on robustness of conclusions to potential sources of bias, among many others. The contribution and benefit of using statistical reasoning to navigate these choppy waters was put sharply into focus, giving statisticians additional responsibilities and at the same time challenging our ways of working.

In this address, I will first look back and draw out some of the highlights of our collective engagement in the pandemic, the challenges and difficulties faced, and give a personal view of how these in turn are giving impetus to fruitful new directions in the merging of statistical principles with constraints of agility, responsiveness and societal responsibilities. What the pandemic has clearly highlighted is how the ubiquitous nature of statistical thinking gives us the capacity to understand and model complex contexts and dynamics. Echoing David Hand's (2009) address, it is the synergy between generic principles and a wide range of substantive problems that makes our discipline so exciting and absolutely core for interacting with a changing and increasingly uncertain world, where much remains only partially known and offers fertile ground for new understanding. Building on this theme, I will turn to present the strategic thinking that, as a society, we are forging regarding our engagement in data science and conclude by reflecting on the common principles that guide our mission and our role in the future.

2 | STATISTICS MAKING AN IMPACT/RULES OF ENGAGEMENT

As beautifully expressed by John Pullinger (2013) in his 2013 address: ‘Statistics is about matters of the highest importance in human affairs’. We are a recognised organisation speaking for the whole of statistics, and throughout our rich history we have endeavoured to intervene and offer our expertise when important statistical issues impacting society at large were not fully recognised by government or the public at large. The 2020 SARS-CoV-2 pandemic is beyond doubt a matter of highest societal importance. What became immediately apparent is that the pace and extent of the threat would place enormous time constraints on the speed required of any response, influencing the nature of statistical endeavours.

The RSS long-term strategic plan has four main objectives, embracing the rich diversity of interest and skills of its members

i for statistics to be used effectively in the public interest;
ii for statistics as a discipline to thrive;
iii for a strong body of professional statisticians;
iv for society to be more statistically literate.

It is the adherence to—and mutual reinforcement between—these objectives within our profession which has guided how we have involved ourselves as statisticians in the societal crisis that the world has experienced since the start of the pandemic. From March 2020, the Society’s response to the SARS-CoV-2 pandemic has been vigorous and nimble, adapting to the urgency while also aiming to keep sight of the medium—and long-term—perspective, making recommendations to address some of the systemic problems in data collection, preparedness and communication that have come to light, partly summarised in the RSS Covid ‘Lessons learned memo’ (2021), and further discussed in a series of evidence sessions that have taken place between April and July 2022.

Our long tradition of engagement has been eminently discussed in previous presidential addresses, (Curnow, 1999; Holt, 2008; Kingman, 1989; Moser, 1980; Pullinger, 2013; Smith, 1996), and reading their distinctive angles but complementary points of view has greatly helped shaped my thoughts.

Contrary to the disenchanted remarks made by Sir Claus Moser (1980), ‘as statisticians, we quickly learn of the crosses we have to bear. We know how people move away from us at parties when they learn of our profession, we know that look of incredulity, amusement and resigned boredom’, since the beginning of the pandemic, statisticians have become sought after guests, not at parties … but on expert panels and in the media. The voice of Sir David Spiegelhalter has now become familiar to many beyond the statistical world and his unique gift in unravelling the mysteries of data and in translating arduous statistical ideas into plain speech and illustrative examples has illuminated the whole country and inspired us all. The awareness that statistics, their production and their interpretation have a direct impact on everyone’s life has been heightened in the public consciousness by the Covid-19 existential threat. In turn, this has conferred important and acknowledged responsibilities on all of us.

Referring to the need for all statisticians to act both competently and honestly when they analyse data and proffer advice to the outside world, Sir John Kingman (1989) says '*It is only if that obligation is fully accepted that they can go on to argue that the world must take our advice seriously*'. The International Statistical Institute code of ethics (2010) states 12 all-encompassing principles, among which

a clarifying obligations and roles, statisticians should take care to stay within their area of competence, and to seek advice, as appropriate, from others with the relevant expertise;
b assessing alternatives impartially, assessing the respective merits and limitations of alternatives;
c maintaining confidence in statistics, alerting potential users of the results to the limits of their reliability and applicability;

are particularly salient for how we have engaged—and will continue to do so whenever necessary—in the health crisis created by the pandemic.

Integral to our capacity for impact are three main pillars, namely our competence, our specific position as independent experts and our communication skills. Key ingredients for success are our willingness to engage, our statistically principled reasoning and our strategy for focusing on issues that fall squarely in our statistical lap. These ingredients are necessary but not always sufficient despite a prolonged effort by the Society since its very beginning to increase awareness of the need for evidence-based decisions. By asking 'what should we be doing to improve and extend our various forms of 'outreach' and to achieve greater influence on the word around us' Sir Adrian Smith (1996) reminded us of the challenges facing statisticians determined to make an impact.

The Society engagement and outreach in the Covid-19 pandemic was substantial. In my discussion of what was achieved and the challenging issues that were facing us, I will not aim at being comprehensive, an impossible task, but will draw on my own personal experience to structure these reflexions.

3 | REFERRING TO STATISTICAL PRINCIPLES: OUR SHARED MODUS OPERANDI

Most of us were not specialists in the statistical analysis of infectious diseases and epidemics, but we all shared our knowledge of statistical principles and our experience of implementing these in our own domain. Our common statistical culture was exercised time and time again by all, to endow us with a critical eye on the value of diverse sources of data and on the different ways evidence was derived and presented to support decisions, and to equip us to be constructive and propose improvements.

3.1 | Questioning the 'value' of different data sets and contributing to collect informative data

In view of the pace of the pandemic and the unpreparedness of surveillance systems faced with a new viral disease, it was particularly difficult at the start to get a clear picture of the extent of the epidemic, the characteristics of the hospitalised patients and the fatality rates amongst the infected. The public was bombarded with a mass of daily numbers. One of our first tasks as statisticians was to unravel their interpretation in the light of our statistical 'savoir faire'. Data never exists without relation to something, and the method used to collect data needs to be scrutinised through a statistical magnifying glass, with close attention paid to the nature, representative or not, of the selection process by which the data is collected. As Sir John Kingman stated: 'Statistical theory is in part a systematic exploration of the ways in which different forms of data can contain [useful information about the questions under consideration]. It should therefore be a guide to the value of different data sets'.

There were many instances where our role was to question and put into perspective the value of different data sets, highlighting potential limitations, difficulties and sometimes completely wrong interpretation. Let us consider the example of the number of daily Covid-19 cases reported by each country, one of the key indicators on the state of the pandemic worldwide that has been routinely included in international data dashboards (e.g., Johns Hopkins University Coronavirus Resource Center, n.d.; Our World in Data, 2022), extensively scrutinised and used for international comparisons. Differences in types of tests used, testing recommendations and protocols, testing capacity and effort, and reporting delays led to artefactual sources of variability, which plague comparative analyses. These were abundantly pointed out by the statistical and epidemiology community (e.g., Ellenberg & Morris, 2021; Spiegelhalter & Masters, 2021). Similar issues of comparability arise with respect to the counting of deaths linked to Covid-19. Numbers quoted originate either from reporting systems used for surveillance linking positive tests for SARS-CoV-2 to subsequent death within an arbitrary chosen time interval, or from the official registration of death certificates. The latter can identify Covid as the underlying cause of death, or include a mention of Covid, and are subject to reporting delays. Finally, the synthetic estimated quantity of excess deaths, which captures direct and indirect effects, requires a careful and explicit choice of baseline for its estimation. The different interpretation of these metrics and some of subtleties underlying their calculations were difficult to communicate to the wider public but constituted an important statistical task.

In the United Kingdom, Covid-19 testing facilities, using a technology considered as gold standard and based on Polymerase Chain Reaction (PCR), were scaled up in the late spring 2020 and their use was targeted towards people exhibiting symptoms and their close contacts. As part of the Department of Health and Social Care (DHSC) Test & Trace system (T&T), later integrated into the UK Health Security Agency (UKHSA), the number of cases,[2] indexed either by specimen date or on report date, as well as the total number of virus tests conducted, were made available daily on a UK government dashboard, which expanded and improved its content progressively. As the epidemic unfolded, it became clear that dashboards were highly effective means of transparently sharing information both visually and through linked downloadable files. Besides being an essential tool for analysts, their use throughout the world has greatly enhanced the public familiarity with data. The design of live and adaptable dashboards with a clear description of content ought to be part of future emergency planning worldwide.

For statisticians, it was immediately clear that a testing protocol targeted at those of increased risk of being infected might be an efficient use of resources but is a likely source of ascertainment bias, if one wants to use these data to estimate prevalence. As a causal link between symptoms and swab testing has been created (Nicholson, Lehmann, et al., 2022a), raw prevalence estimates derived from such data (as a proportion of the tested population) will tend to have an upwards bias, which can be furthermore influenced by variations in testing capacity.

The fundamental concept of random sampling, by which the link between symptoms and testing can be broken, led early in the United Kingdom to the establishment of two randomly sampled population surveys of Covid-19 prevalence, the Office of National Statistics (ONS) Covid-19 infection survey and the REal Time Assessment of Community Transmission (REACT) survey. Both surveys are designed to be representative of the UK population and have been invaluable as a reliable source of unbiassed information for estimating absolute and changes in SARS-CoV-2 prevalence since their inception in 2020. To my knowledge, such a sustained roll out over nearly 2 years of prevalence surveys based on random sampling has not been replicated in other countries. It exemplifies the benefits of having in the United Kingdom an effective and recognised national statistics system and a strong statistical tradition, supported by an active engagement of statisticians at large. As the epidemic entered a new phase in 2022 in the United Kingdom with wide vaccination coverage and dominance of less severe variants, these surveillance studies were reduced in size or terminated from Spring 22. Going forward, it will be important for our community to engage in a broad statistical conversation on different forms of active surveillance, their cost, benefits and operational viability, so that lessons can be learned and the best means of ramping up active surveillance are debated, a conversation that the RSS hopes to facilitate.

3.2 | Adding value through evidence synthesis

Coming back to John Kingman's notion of 'value of different data sets', statistical principles also give us a guide on how to perform data synthesis between different sources of data, a task which is critical when adopting a system's view of disease surveillance. To carry out such data synthesis while ensuring adequate propagation of uncertainty, one may turn, for example, to the generic framework of Bayesian graphical models. In the case of Covid-19 prevalence estimation, it is natural to surmise that 'added value' can be gained by combining the information provided by surveys based on random sampling with that encapsulated in the large, spatially granular but biassed testing data recorded by Test & Trace. In a project[3] undertaken in partnership with UKHSA, fusion of the routinely reported testing data with information from the REACT study was implemented using a Bayesian data synthesis framework, providing a weekly debiassed estimate of the true prevalence at a local scale (Nicholson, Lehmann, et al., 2022a). This debiassed estimate is regionally compatible with the survey estimate from random sampling but more precise, as expected from Bayesian principles (Figure 1). As a by-product of this model, the ascertainment bias of the testing data is quantified. Such a debiassed estimate was made available to the UKHSA from March 2021 onwards.

In general, faced with diverse sources of data that can be linked to common latent quantities of interest, we can be guided as to their 'value' by carefully understanding and modelling the data collection process. If appropriate, one can proceed to perform a principled data synthesis to gain precision. Such approaches have been used with success in the past for reconstructing prevalence of other infectious diseases, such as the impactful Multi-parameter Evidence Synthesis framework for HIV prevalence estimation developed by Goubar et al. (2008). However, there remain substantial challenges in operationalising at pace a synthesis of diverse sources of data potentially useful for surveillance; these are challenges that in turn create opportunities for developing enabling methodology to enhance preparedness and upscaled delivery when needed in emergency (see Section 4.3).

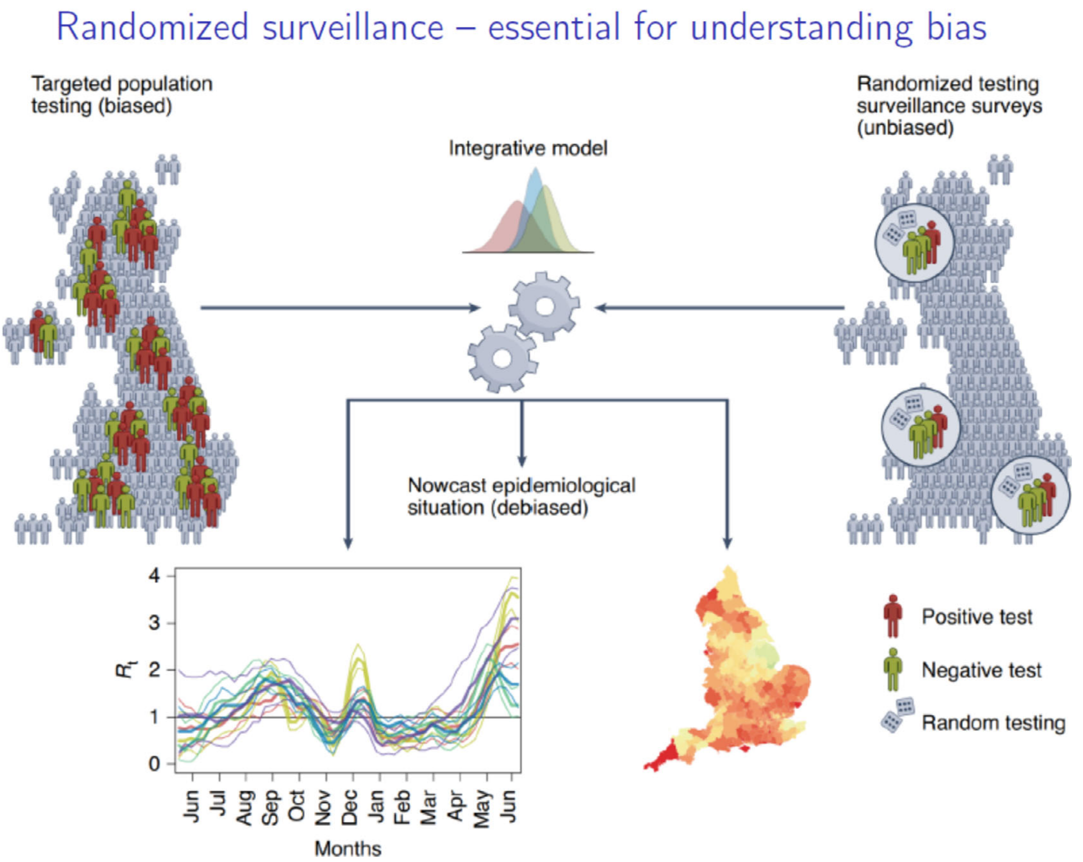

FIGURE 1 Nowcasting the spread of SARS-CoV-2 by combining randomised testing surveillance data and finer-scale spatiotemporal targeted testing data. Reproduced from Rossman and Segal (2022) commentary on Nicholson, Lehmann, et al. (2022a)

3.3 | Evaluations of treatments, systems and policies

A bitter lesson that governments have learnt is that a pandemic moves fast, and that public health and other measures need to anticipate the intrinsic exponential growth of the transmission process in a fully susceptible population. The good news of the deployment of the first Covid-19 vaccines in January 2021 was accompanied by the bad news that a new 'alpha' variant of the virus had emerged in the last weeks of 2020, quickening the disease's spread, and pushing up the number of patients treated in hospital due to increased severity. As I was starting to write this address at the beginning of 2022, yet another variant, omicron, has become dominant with different transmission and severity characteristics. Faced with such a substantial threat to the health of populations in the winter 2020, the United Kingdom and other governments had to intervene swiftly and continually adapt their response, even when evidence was limited and uncertain. But the urgency of any situation should not deter from planning in parallel smart evaluation strategies and the collection of evidence on the effectiveness of the arsenal of therapeutic and public health policies that can be deployed. We need to strive to learn at the same time as acting, a challenging task. Randomised policy evaluations are often curtailed by ethical and operational constraints which restrict the space of possible designs, while policy evaluations using an ecological-type design are subject to confounding and different types of bias.

3.3.1 | Platform trials

In the clinical sphere, a shining example of 'delivery in real time' of answers on the efficacy of different lines of treatment for hospitalised Covid-19 patients is that of the large-scale RECOVERY trial. In less than 4 months, RECOVERY led to its first conclusive evidence on improved treatment options for severe patients. The United Kingdom has a stellar record of designing and implementing randomised trials, starting back in 1948 with testing the use of streptomycin for treating tuberculosis (Medical Research Council [MRC], 1948). Building on advances in trial design methodology, the statistical team designing the protocol of RECOVERY proposed an adaptive multi-arm platform design, which was quickly approved in full recognition of the urgency of the pandemic threat; new treatments could be added, and useless ones dropped quickly. One lesson can be drawn from this exemplary story. The key components of success were the conjunction of (i) a well-organised and collegial trial design community, where challenging and desirable developments in statistical methodology are continually identified and tackled, (ii) a well-established regulatory framework which was responsive to the time pressure, and (iii) the existence of a unified national health care system which enabled sharing of protocols and efficient trial recruitment.

These components of success are not easy to replicate in the wider context of rollouts of new surveillance systems or policies. In this domain, there is no well-defined regulatory framework to follow, and the evaluation of impact of a new policy or system is usually considered on a piecemeal basis. If we want to make progress, improve preparedness, and increase accountability, it is important that statisticians contribute to bring about an increased recognition that monitoring a new system or considering how to evaluate a policy is best thought of in an interdisciplinary way, to be put into place right from the start of the policy or new system, even when action is urgent.

3.3.2 | Performance of the T&T

The T&T was set up in May 2020. Despite the government spending a vast amount on testing, tracing and isolating schemes, there was no routine reporting of simple performance indicators informative on its ability to break chains of transmission, such as the time taken to reach a contact during the likely window of infectiousness of an index case. The RSS Covid-19 Task Force issued a statement on the 23 July 2020, with several recommendations on 'how efficient statistical methods can glean greater intelligence from T&T'. RSS recommendations included the use of record linkage between index cases and their contacts who subsequently took a swab test, and embedding a study based on random sampling of households to investigate secondary attack rate and compliance with isolation and requests to stay at home. Such a linkage was subsequently performed by DHSC T&T, as can be seen, for example, in Lee et al. (2022), but regrettably not used as the basis for regular and transparent reports, leaving the public without interpretable information to assess the performance of the T&T system.

In September 2021, more than 16 months after its start, DHSC T&T produced an updated modelling-based analysis of the impact of the T&T system on transmission reduction (the Canna Model) by comparing it to a counterfactual scenario. Linkage within the Contract Tracing and Advisory System and estimates of the secondary attack rate and of the probability of compliance with isolation recommendations were crucial elements feeding into this analysis. Echoing some of the RSS recommendations, a prospective cohort study, the ATACC study, was set up in September 2020, to study community transmission and its evolution during the pandemic, study which published its findings only in early 2022 (Singanayagam et al., 2022). These reports provided useful information, but their late timing did not allow any conclusions to be fed into any system improvement for breaking chains of transmission. Again, one sees the importance of timeliness in delivery, which challenges the way we do design and analysis.

Thinking about means of evaluations and engaging statisticians in this conversation right from the start of a new surveillance system is a needed change of culture that will enable improvements in real time and will increase public confidence in the usefulness of public health policies.

3.3.3 | Controlling infections in schools

Setting up randomised trials to evaluate—in context—the effectiveness of recommended transmission control policies is the agreed route to solid evidence. But targeting the evaluation on relevant policies is problematic in a fast-moving pandemic. Infection control in schools is a good example where a delicate balance is required between the need for evaluation and the pace of decision making. Keeping children in schools has been a major concern of governments throughout the pandemic as school closure is damaging in many ways. The challenge for decision makers lies in balancing the health risks from infection of children in schools and potential secondary infections in their families and wider contact networks, with the risks of loss of both academic and social skills for the young and increased inequality, as well as the economic and social impact of parents having to stay at home (Royal Society DELVE Initiative, 2020). Hence infection control in schools has given rise to heated debates on viable alternatives to the testing and isolation strategies recommended for the general population, which tend to exclude many children from face-to-face schooling.

The availability of rapid testing using Lateral Flow Devices (LFD) opened the possibility of introducing new infection-control policies in schools combining rapid testing and modified isolation guidelines, which could potentially increase face-to-face attendance. After the reopening of schools in March 2021, an open-label cluster-randomised trial including over 200 schools was set up to compare the effectiveness of the standard policy of self-isolation of school-based Covid-19 contacts for 10 days to a new control policy involving voluntary daily testing by LFD of contacts for 7 days while allowing LFD negative contacts to remain in school. This trial reported in October 2021 and found that daily contact testing was non-inferior to self-isolation (Young et al., 2021). But the public health active surveillance context had changed again, and LFD testing was starting to be made available to schools for asymptomatic testing, an option not considered in the trial. In a fast-moving context such as a pandemic, we are thus faced with the need to increase the pace of our well-trodden routes of evaluation through randomised trials, otherwise we run the risk of limiting the usefulness of the collected evidence.

4 | CHALLENGES

4.1 | Communication

Two of our strategic objectives: (i) for statistics to be used effectively in the public interest and (ii) for society to be more statistically literate, have been put under a massive stress test during the pandemic. Communicating clearly to a wide public how to interpret or ignore the daily headlines that were hand-picked by the media from a vast quantity of reported data was challenging but essential in view of the heightened public interest in statistics. Already in May 2020, David Spiegelhalter was denouncing in a widely watched television programme the 'number theatre' of meaningless league tables across countries that were alluded to during Number 10 press conferences. From the start, the wide RSS membership, the statistical ambassadors and the Covid-19 Task Force engaged in sustained efforts of communication via the Science Media Centre or directly with journalists, the BBC and other channels. Through the RSS voice, statistical concerns were openly brought to the public attention; for example, by explaining the influence of reporting delays in death certification, by calling for increased transparency and release of data and evidence supporting policy decisions, by reminding of the importance of designed evaluation studies based on random sampling, and by emphasising that false positive, false negative, background prevalence and context of use need to be considered when planning mass testing exercises. Starting from the RSS FAQ and continuing through a fortnightly Observer column, David Spiegelhalter and Anthony Masters explained to a keen public many of the issues behind the numbers that were regularly quoted in the press and often misinterpreted, culminating in the publication of their popular book 'Covid by Numbers – Making Sense of the Pandemic with Data' in November 2021 (Spiegelhalter & Masters, 2021).

The involvement of prominent statisticians with the media has been critical in raising the profile of statistical understanding in the general population and the pandemic has shown how important this is. It has also shown that communicating clearly requires finely tuned skills and cannot be improvised. The five rules for evidence communication outlined in Blastland et al. (2020), 'inform, not persuade; offer balance, not false balance; disclose uncertainties; state evidence quality; and inoculate against misinformation', are excellent guidelines, but unfortunately all too easy to deviate from without a good dose of practice when questioned on live radio or TV! Learning to champion impartial evidence in an accessible way is a challenge that future generations of statisticians should engage in actively and enjoy.

One of the most delicate elements of our task is how to communicate uncertainty so that trust in the statistical results is increased and not tarnished. To echo the aphorism '*La vraie science est une ignorance qui se sait*' attributed to Michel de Montaigne, a French philosopher, true science acknowledges the limit of what we know. As eloquently discussed by Spiegelhalter (2017) in his presidential address: 'we need to encourage stories that are true to the evidence, its strength, weaknesses and its uncertainties'. We also have a duty to speak out when faced with false statements or analyses, something that Spiegelhalter aptly coined 'muscular uncertainty'.

In the specific context of a pandemic, quantifying transmission dynamics, for example by estimating headline summaries such as the effective reproduction number R_t (see Pellis et al., 2022) was of key interest to public health officials. At some point, such summaries were discussed daily. Communicating structural and statistical uncertainty associated with the complex epidemic modelling work that was undertaken by different teams to quantify R_t was an arduous task, prone to misinterpretation. It is nevertheless important that open scrutiny of the plausibility of assumptions underlying transmission models is encouraged, as well as a non-technical discussion of the limitations of the data sources used to estimate key latent quantities. This will increase informed public debate and counteract deliberate misinformation.

In addition to our familiar challenge of communicating uncertainty appropriately, scientists and statisticians alike were faced with another difficult task, that of communicating this in an infectious disease context, a context which creates dependent chains of events. When the relation between individual and collective risks is non-additive, each individual decision affects all. The interactive Covid-19 transmission routes graph produced by the Winton Centre in collaboration with the British Medical Journal is a creative proposal in that direction. Going further, illustrating the collective impact of individual decisions in a non-additive risk context, and communicating this to a wide audience, through graphical scenarios or other means, are important tasks. These are needed to increase public understanding of the benefits and difficulties of adopting the necessary system view for such society-wide issues. Explaining collective risk would be best tackled from multidisciplinary collaborative efforts including social and environmental scientists as well as statisticians and health experts.

Besides communicating to the public, it is also important to foster good practice in the media. By establishing yearly RSS awards for statistical excellence in journalism since 2007, the RSS has been proactive in raising awareness and understanding in the media of what statistics are, what they can be used for, and of the need for integrity in the explanation, use of statistics and avoidance of distortion. In its Covid 'lessons learned memo', the RSS called for the media to step up to its responsibilities regretting that some media outlets chose political lobbyists rather than science journalists to cover the government briefings, and praising institutions like the Science Media Centre, Full Fact and Our World in Data. Our challenge will be to build on our experience of communication during the pandemic to support the good presentation of evidence on all future important societal issues.

4.2 | Agility in policy evaluations

Faced with a fast-moving situation where, for example, new variants with different transmissibility and vaccine escape properties emerge in an unpredictable fashion, how to be as agile as possible in targeting policy trials to provide timely evidence is a challenging question. I would argue that there is a role for integrating this process with additional insights provided, for example, by modelling or simulation exercises. Instead of opposing modelling scenarios and designed evaluation studies, there is much to be gained from investigating how these different perspectives can be mutually reinforcing. Individual level simulations, using for example agent-based models, can be easily tailored to reproduce specific contexts of transmission (schools, specific types of workplaces, transport systems etc.) and designed to mimic the effect of a wide array of control policies under consideration. By their nature, such simulations are eminently flexible, can be programmed efficiently and deployed quickly. They can be used to gain qualitative and quantitative understanding on key drivers behind the performance of policies. The information they provide could be dynamically fed into the design of policy evaluation trials, potentially creating new or modified arms, which take into account changes in knowledge and circumstances, in the spirit of adaptive trial methodology. Of course, this would be most relevant in situations where there is already substantive knowledge of the specific context to anchor the assumptions underlying the generative simulation models.

In the previously discussed school context, key simulation parameters relate to class size and pupil contact structure, infectivity curves over time, probability of transmission following an infected contact, operational characteristics of rapid LFD tests to be used, testing protocols (e.g., including regular asymptomatic testing or not) and isolation policies. Simulation approaches can then be used to gain further understanding of the influence of key parameters, rank policies and investigate robustness of this ranking to changes in parameters, providing ‘robust candidates’ to be taken forward for evaluation through statistically designed study. New information emerging from empirical or other studies can be speedily incorporated into the simulations to inform the need for any change in the design. Initial work along this line, both in the United Kingdom and in France, delivered new quantitative understanding on effectiveness of school policy for controlling transmission of SARS-CoV-2, highlighting the benefit of regular asymptomatic testing in schools as a robust and effective policy across a range of simulation parameters (Di Domenico et al., 2021; Kunzmann et al., 2022; Leng et al., 2022).

In summary, there has been an unprecedented need for evaluations of treatments and infection control policies at pace. While statistical methodology for adaptive trials was already mature, the Covid-19 pandemic has shown that there remain considerable challenges in deploying purposely designed evaluations of policies in real time. Developing adaptive strategies for merging simulation-based and in-context designed evaluations is an intriguing and promising avenue to increase flexibility in the domain of policy evaluations. In this paragraph, I have focused on discussing agility of policy evaluations within the context of studies with designed sampling strategies. I have not attempted to cover the area of policy evaluations using observational data, which draws on analysis tools belonging to the research-active framework of causal inference (Hernán & Robins, 2020). Recent developments in causal inference methods for combining clinical trials with external observation from observational studies (see the review in Shi et al., 2022) might provide another interesting route for increasing adaptability of policy trials.

There is still much that creative statistical methodology can contribute to the domain of policy evaluation, an important area of our mission encapsulated in our strapline ‘Data, evidence, decisions’. Our acquired experience during the pandemic is a good platform for pursuing new research to advance the embedding of data science approaches into policy evaluation.

4.3 | The role of statistician as an expert at the interface between statistics and decision-making

Scientific advice from experts in a wide range of disciplines and domains has been the backbone of governments’ answers to the pandemic in many countries. Public health institutions throughout the world have naturally been the first to engage in the battle, but many countries have also created de novo scientific advisory structures to mobilise a mix of expertise adapted to the crisis. In the United Kingdom, pandemic emergency preparedness has at its core SAGE, the Scientific Advisory Group for Emergencies, a number of domain specific expert subgroups, and the overarching high-level advice provided by the Chief Scientific Officer and the Chief Medical Officer. There are clearly dangers for scientific experts to be perceived as closely associated with government or industry. The current crisis is no exception, with some scientists at the receiving end of media insults (Nogrady, 2021).

The distinction between presenting scientific evidence and giving an opinion on public health policies adopted by governments is easily blurred. How many times have we heard since the beginning of the pandemic the sentence ‘guided by scientific advice’ to precede a government announcement on new public health recommendations or legally binding rules designed to control spread of the SARS-CoV-2 19 virus? But how the evidence gathered by experts is weighed and translated into actions is rarely formalised. Once appraised of evidence and associated uncertainties, it is governments which balance competing societal needs, political priorities and economic outlook when choosing a set of policies. The ultimate policy chosen is itself a political decision and outside the realm of statistics. As statisticians and data scientists, we must add our voice to clarify the distinction between scientific evidence and its consequences. We need to highlight where our expertise should be called upon and when it is not our role to engage and help policy makers understand the ways in which current and often time limited evidence can inform their actions.

As argued by Robert Curnow (1999), statistical expertise has a unique place in the sequence of events that start with scientific findings and end with a specific policy implementation. Along this process, many tasks fall squarely in the statistician's corner: checking or advising on the analysis and interpretation of the data which support scientific conclusions, identifying the need for new data to fill knowledge gaps and helping in designing data collection, highlighting sources of uncertainties and quantifying their impact whenever possible, investigating sensitivity to assumptions, and assisting in evidence synthesis.

Clearly, time is of the essence in a fast-moving pandemic and the statistical tasks listed above must be performed as best as possible given the constraints of delivering in a timely manner. Never have the rules of statistical thinking outlined by John Pullinger (2013) been more relevant, namely:

- Statistics are provisional: Findings can always be improved. We must be willing to recognise the validity of the figures that are available today but ready to applaud rather than criticise when tomorrow there is better information available through which to revise them
- Statistics are uncertain. We should be sceptical of peddling impossible guarantees, rather than demanding them, and celebrate those who tell us about risk and imprecision.

It is not easy to capture in full the contribution of statistical expertise to the public health decisions made during the pandemic. One major route was through involvement in SPI-M, a sub-group of SAGE, pre-established to respond to severe threats from new influenza strains. Statistical inference and data were included into the SPI-M epidemic models in a variety of ways, at minimum to compute plug-in estimates for key model parameters, or comprehensively to implement a full joint model of different surveillance data sources within a Bayesian framework such as in the PHE-Cambridge model (Birrell et al., 2021). The output from a set of models implemented in SPI-M was summarised through a meta-analytic approach, leading to reporting ranges for estimates of the growth rate and the effective reproduction number, key indicators fed into SAGE as the pandemic evolved and new variants became predominant. Besides epidemic modelling, a range of other statistical analyses were carried out by SPI-M teams, such as assessing the severity and risk of hospital admission for new variants (Nyberg et al., 2021, 2022), providing crucial statistical evidence for public health and NHS planning.

In addition to the work of SPI-M, other data-driven initiatives got under way. In April 2020, the Royal Society convened the Data Evaluation and Learning for Viral Epidemics (DELVE), a multi-disciplinary initiative aimed at supporting a data-driven approach to learn from the different approaches that countries were taking to manage the pandemic. It put data at its core and gathered experts across data sciences, epidemiology, public health, infectious diseases, economics, behavioural sciences, immunology and more. Its focus was to write timely reports related to policy issues that were under discussion at governmental level, such as mask wearing, test and trace systems, school policy, and to present these at SAGE. On the RSS front, besides the advocacy activities of the RSS Covid-19 Task Force and their indirect impact, a Turing-RSS Health Data Lab, led by Chris Holmes and supported on the RSS side by Peter Diggle and me, was established in October 2020 in partnership with DHSC and the Joint Biosecurity Centre (JBC). In October 2021, the Health Data Lab was embedded in UKHSA until its closure at the end of June 2022. The strategic vision of the Lab was to support the UKHSA through an embedded data science collaboration, working towards the development of an interoperable framework that would provide quantitative evidence to decision-makers in an agile and responsive fashion and ultimately deliver tools of generic long-term value for health surveillance.

From a strategic perspective, it is valuable to reflect on the last 2 years and extract a clear message on how statisticians and data scientists could and should be involved as experts when society is confronted by a situation of emergency. To kick start this reflection process, an RSS debate on evidence and policy was organised in June 2022 as part of a series of RSS events on statistical aspects of the UK's response to Covid-19, ahead of the public inquiry. One important question debated was how we could build on our experience to prepare and train a reservoir of statisticians and data scientists for such emergency situations, and what role could the RSS play in helping this process. As a voice for statistics, the RSS should continue to promote the recognition by policy makers of the usefulness of including a statistical perspective right from the start on many of the issues that would confront government in an emergency. Drawing on their experience in DELVE, Lawrence and Montgomery (2020) advocated the need for a multidisciplinary initiative to build a community of researchers from data science and other disciplines who are able to rapidly mobilise and respond to future policy challenges. This is an interesting avenue to pursue to generate a change of culture and broaden the contribution of data science to policy-making.

4.4 | The road to the long-term impact of recommendations

During the pandemic, the use of different types of Covid-19 diagnostics tests was a recurrent topic of debate which touched on many statistical issues. To clarify these issues and create the basis of a dialogue with the responsible health authorities, a comprehensive statistical appraisal of *Diagnostic Tests* was undertaken by an RSS Working Group (rss.org, 2021). Before discussing the generic recommendations outlined in the working group's report, I would like to turn to the past, and take as a point of reference the influential work started in the 1990s that the RSS produced on ranking of performance indicators, so-called league tables. In choosing to refer back to this example, I am conscious that other striking ones will come immediately to readers' minds, an excellent state of affairs! Here my purpose is not to be exhaustive but to draw out some general threads.

The RSS has made far-reaching impact on highlighting the statistical issues related to monitoring the performance of public services, such as schools and the health service. This was due to a fortuitous conjunction of statistical issues related to league tables coming into focus in parallel to the development of data bases on public services and increasing demand for transparent accountability of public services and professionals. The topic was kicked off by Goldstein and Spiegelhalter's (1996) paper, which discussed the limitation of published league tables, making a strong case for considering uncertainty of these rankings. It was then sharply brought into public focus by two official enquiries, the public enquiry into paediatric cardiac surgery at the Bristol Royal Infirmary, and the Shipman enquiry. Biostatisticians were officially approached and commissioned to re-analyse some of the data (Spiegelhalter et al., 2002) or to assess the feasibility of routine monitoring (Aylin et al., 2003). Besides reporting carefully their findings, including when the data were not strong enough to lead to a firm conclusion, the analytical work carried out led to a discussion of broad-ranging issues regarding the information system and data collection needed to be in place for conducting routine monitoring. It also fostered reflections on appropriate statistical tools for performance monitoring (Spiegelhalter & Best, 2004). The RSS then convened a working party on *Performance Monitoring for the Public Services*, which gave comprehensive recommendations in a milestone widely cited paper (Bird et al., 2005), including the need for setting out detailed protocols and specifying objectives, having independent scrutiny, reporting uncertainty, including considerations of ethics and making efforts in explaining how to interpret performance monitoring indicators to non-statisticians. During that period, continuous monitoring of health care data, for example to detect unusual patterns of patient mortality within General Practices, was put into practice through the establishment of the Doctor Foster Unit at Imperial College in 2000.

Looking back to this example, we see that the strong engagement of statisticians in performance monitoring impacted not only important public enquiries but established a generic framework and statistical tools to provide long term guidance, which has become a standard reference. I would now like to draw a parallel with the engagement of statisticians and the RSS on diagnostic tests for infection by SARS-CoV-2.

Early in the pandemic, it became clear that one instrument essential to all governments managing the pandemic was to have available a reliable in vitro test to detect if a person was currently infected by the SARS-CoV-2 virus. Thanks to the sharing of the SARS-CoV-2 genetic sequence by Chinese scientists on 10 January 2020, an RT-PCR assay was published on 23 January 2020 (Corman et al., 2020); this could confirm the presence of the SARS-Cov-2 virus strain in a person's nasal or throat DNA samples using sequence amplification. Note that RT-PCR sequencing techniques require processing in a laboratory and preclude return of a rapid, on the spot, test result. Many countries quickly embarked on expanding genetic sequencing capacity to deliver such RT-PCR SARS-CoV-2 tests on a large scale as well as building pipelines to centralise and report results in a timely manner.

While the ramping-up of massive sequencing efforts was taking place, research by the private sector on alternative rapid antigenic or antibody tests that did not require specialised equipment and could be processed and read by non-specialist staff and the wider public, got under way. The concept of rapid testing was attractive both to the public and to governments, but the statistical community became quickly aware that, alongside the biotechnical challenges that needed to be overcome, many issues of a statistical nature had to be aired. The RSS Covid-19 task force, supported by RSS fellows expert in diagnostic tests, set out to inform and clarify some of these issues. A first distinction concerns the intended purpose of a test—whether it was for population surveillance or for clinical diagnosis—as the specification of thresholds for the operational characteristics should take this into account. In parallel with criteria released by the Medicines and Healthcare products Regulatory Agency (MRHA) on point-of-care tests for SARS-CoV-2 antibodies, the task force issued a statement in April 2020 explaining that surveillance studies might use tests with lower criteria on clinical sensitivity and specificity. As the government continued to evolve its strategy on approval and use of diagnostic tests, the task force repeatedly made the case for transparency of data on the performance of diagnostic tests, and the importance of carrying out context-specific performance evaluations besides ones in 'idealised' laboratory-controlled settings.

Against this background of debate about standards applied to the evaluation of in vitro diagnostics tests, it became clear that there was a need for a comprehensive review of the statistical evidence needed to assure the performance of new tests, for patients, decision makers and regulators. An RSS Working Group on Diagnostic Tests was convened in the summer of

2020, co-chaired by Deborah Ashby and Jon Deeks, which made important recommendations on standards for study design and presentation of evidence, planning for future pandemics and regulations. The milestone report published in June 2021 offered to a wide community of users and decision-makers a thoughtful decoding guide on the many subtle statistical issues related to diagnostic testing and created a platform for further interactions with regulators currently being pursued.

In summary, similarly to the case of performance monitoring, the range of statistical issues related to the approval and use of new diagnostic tests were elucidated and brought to the attention of a wide range of stakeholders. In vitro diagnostic tests are an integral part of the future of health surveillance, and the recommendations of the RSS provide an important blueprint for their best use for the benefit of patients, the general public and policy-makers.

5 | WHAT IS THE STATISTICAL LEGACY FROM THE PANDEMIC?

Throughout its history, advances in statistics have arisen from demands in science and in society. A legitimate interrogation is to ask what new directions in statistics may have resulted from the SARS-CoV-2 pandemic? Before Covid-19, the last global infectious disease outbreak was the HIV/AIDS epidemic. This created demands on the statistical analysis of infectious diseases which led to advances in survival analysis methodology and inference for partially observed Markov processes. By contrasting the tasks facing statisticians in these two pandemics, a discussion started by Ellenberg and Morris (2021), we can reflect how the statistical concepts and methods built for the HIV pandemic have informed our statistical handling of the current pandemic and tease out some of the new directions created by the specific context of the SARS-Cov-2 pandemic.

5.1 | A retrospective look at the statistical influence of the HIV and AIDS epidemic

From its first appearance in the public consciousness in the United States in 1981, the AIDS epidemic developed into a major crisis, which resonated worldwide in the 1980s. AIDS was not perceived as a major threat from the beginning but viewed at first as another infectious disease to which well-tested method of surveillance, research, prevention and treatment could be applied to. This view was disputed by some scientists (Gallo & Montagnier, 1988). It gradually became recognised that despite identification of the HIV virus, understanding of the virus's target in the body and availability of a blood test, the uncertainty on the lengthy time interval for a person infected with HIV to develop a full-blown fatal AIDS syndrome, and the fact that no treatment was available till the first drug approved in 1987, clearly distinguished infection from HIV from previously encountered viral infections. In late 1987, the RSS initiated a series of workshops for statisticians active in mathematical modelling of the AIDS epidemic and scheduled on open discussion meeting (Gore & Armitage, 1988). The president, Sir John Kingman together with James Durbin, Sir David Cox and Michael JR Healy, drafted a statement to capture the statistical requirements of the AIDS epidemic with a list of recommendations on data requirements, data sharing, expansion of testing, information campaigns, patient management and implications for planning of resources (Kingman et al., 1988). Much of the substance of these recommendations has been reiterated by statisticians and the RSS in the current crisis, see Covid 'Lessons learned memo' (2021).

The public health questions posed by the HIV epidemic inspired a rich body of statistical literature which laid down fundamental statistical approaches for estimating the burden of infectious diseases. Estimating the distribution function of key characteristics such as the incubation time or the time between infection and transmission to a sexual partner was not amenable to standard survival analysis because events at either end of the time interval were typically unobserved, a data structure referred to as 'doubly censored' by De Gruttola and Lagakos (1989). This stimulated substantial methodological developments in time to event analysis, in particular of non-parametric methods (Jewell, 1994), developments which have had applicability well beyond the HIV context. Measuring the HIV/AIDS epidemic in terms of prevalence and incidence also presented important statistical challenges which evolved as relevant data sources grew. Thoughts were given to the usefulness of different types of surveys: probability-based nationally representative surveys, or surveys focused on specific groups such as pregnant women or high-risk subpopulations, and to ways of combining these to estimate the underlying prevalence (Brookmeyer, 2010). The impactful technique of *back-calculation*, combining AIDS surveillance data with information on the incubation period to infer historical HIV infection dates and incidence trends, was first proposed during that period (Brookmeyer & Gail, 1988). It is not the place to go into more details. My purpose was simply to highlight the considerable statistical legacy spurred by the HIV epidemic, a legacy which has come to the fore when analysing the natural history of SARS-Co-2 infections.

5.2 | Modelling and inference at pace

In view of their challenging natural history and mode of transmission combined with their huge societal impact, the HIV and the SARS-CoV-2 epidemics stand apart from other health threats from transmissible diseases in the last 40 years. But there is a striking difference of pace. For the SARS-CoV-2 pandemic, the time axis is hugely compressed creating specific expectations on the delivery of statistical analyses. As pointed out by Brookmeyer (2021), the United States surpassed a cumulative Covid-19 death toll of 400,000 within 11 months of the first reported US Covid case, while it took more than 16 years for the US cumulative death toll from AIDS to reach 400,000 from the first reported AIDS case.

Right from the start of the SARS-CoV-2 pandemic, the community of statisticians and modellers working on infectious diseases set out to modify and adapt an array of models and computational tools that had been developed since the AIDS/HIV epidemic in response to a range of emerging pathogens, such as the influenza A (H1N1) pandemic. The interface between statistical analysis and non-linear dynamical systems had considerably grown since HIV and benefited from advances in computational power, availability of data, Bayesian approaches and algorithms for fitting complex systems and machine learning methods (Halloran, 2021).

In the United Kingdom and abroad, this community organised itself on a war-like footing in large teams with complementary skills to produce reviews, situation estimation, short-term predictions or what-if modelling scenarios at pace, which were fed into government advisory groups and scrutinised by public health authorities. Altogether many flavours of epidemic models were developed for tracking Covid-19 disease transmission around the world creating a huge body of literature, which was hard to keep abreast of. Fitting such models in real time to imperfect, incomplete and sometimes conflicting data sources was extremely challenging and often required the development of approximate computations. Usual ways of evaluating model performance by comparing predictions to observed data were confronted to an unstable non-stationary context, rendering their interpretation difficult. The continual back loop between behavioural changes and epidemic evolution rendered all but very short term predictions somewhat futile. Key summary statistics characterising the current state (e.g., prevalence, hospital occupancy) or epidemic trends (e.g., reproduction number, growth rate) of the evolving epidemic were estimated to guide public health policies. The instantaneous reproduction number R_t, measuring the average number of secondary infections generated by a primary case at time t, became one of the key metrics used in the United Kingdom. Opinions vary on the usefulness of this summary measure and the instantaneous epidemic growth rate r_t, defined as the rate of change of the log-transformed case incidence, was used in complement to R_t. To overview the body of modelling work and discuss openly the statistical issues around measuring and estimating transmission and how to operationalise these estimates under severe time pressure, the RSS convened a special topic meeting on Covid-19 transmission, co-chaired by Peter Diggle and me. The set of position papers and associated discussions give an insightful overview on the interpretation and misinterpretation of the different metrics that can be used to characterise the course of an epidemic, the factors influencing these, how sensitive their estimation might be to the form of the model and the noise in the data. The rich output of this meeting decodes for readers the many subtle statistical issues linked, for example, to the choice and relative quality of data streams chosen as input into the models or the effect of heterogeneity. In a complementary fashion the broad statistical issues linked to Covid-19 were also debated in two RSS discussion paper sessions, following a special call for papers on 'statistical aspects of the Covid-19 pandemic'. As we transition from urgent modelling efforts to the management of 'living with Covid' around the world, continuing a reflection on the impact of the intensive modelling work that was produced will be important, with a view to drawing lessons for improved preparedness and communication of the role of epidemiological modelling both to decision-makers and to the public.

5.3 | Statistical inter-operability

Interesting challenges to conventional statistical practice arose during the pandemic from the need to analyse real-time, messy data from diverse sources to answer constantly changing questions from the health authorities. Agility in statistical modelling and analysis became a premium in contrast to conventional statistical analysis protocols that target specific research questions. By its practice developed in partnership with UKHSA, the Turing-RSS Health Data Lab experienced at first hand this need for agility, while being mindful to preserve a coherent treatment of uncertainty and taking an overarching system's view of many connected questions. This led the Lab to articulate the concept of '*statistical interoperability*' as an aspiration for the future design of real-time disease surveillance systems (Nicholson, Blangiardo, et al., 2022b). At its core, statistical interoperability is driven by the need to optimally fulfil the joint goals of *agility* (rapidly communicating and interlinking statistical modelling output across analyses, with modular components transferable across health security problems); *sustainability* (development of a common high-quality open-source analytics code base that grows over time); *transferability* (co-ownership of projects between academia and health security teams); and *preparedness* (solutions developed for one particular problem, e.g. tracking Covid-19, can be quickly re-purposed to meet future public health challenges). In a similar spirit, Yu and Singh (2022) articulate principles for rapid response data science based on their experience of Covid-19 forecasting.

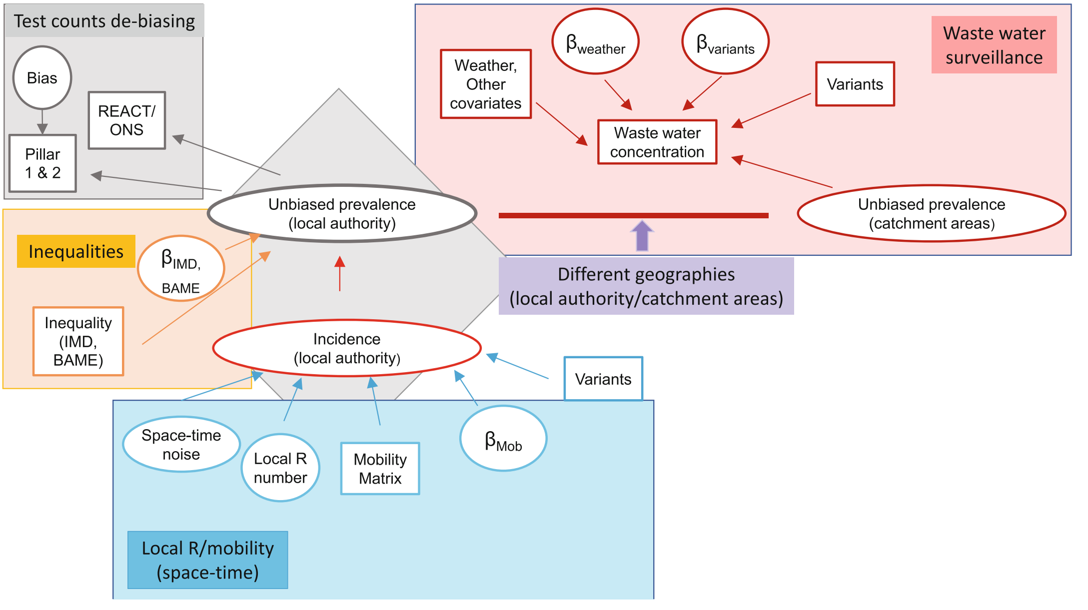

FIGURE 2 Schematic graphical representation of interoperability between projects undertaken in the Turing-RSS health data lab between October 2020 and June 2022. The output of the core project on de-biassing test counts feeds into projects on social inequalities of infection rates, influence of mobility on transmission, and monitoring of wastewater for disease surveillance

The key desirability of modularity leads in a natural way to choosing Bayesian graphical models as core framework for building interoperable model structures, as it clarifies information flows and key assumptions, and allows inclusion of new sources of data in a coherent manner. Essentially, this framework facilitates breaking down a complex model into smaller self-contained modules, that can be developed in parallel and updated when new knowledge becomes available, and then joined when needed in a soft serial fashion, using their shared latent quantities as key linchpins. An illustration of the interoperability principle as implemented by the Health Data Lab to build links between projects is shown in Figure 2.

Delivering interoperable computations with due consideration to uncertainty has relied on an established body of statistical research on modular inference taking its roots in Bayesian melding (Poole & Raftery, 2000), originally proposed to carry out formal inference for deterministic simulation models, an important context in many scientific disciplines including, for example, population dynamics. It builds on work on evidence synthesis (Presanis et al., 2014) and Markov melding (Goudie et al., 2019), a generic framework for forming suitable joint models when joining sub-models, and on semi-modular inference developed to temper the influence of less-reliable data sources (Carmona & Nicholls, 2020). Whatever framework is chosen, I view the impetus to progress our common knowledge and practice of computationally efficient and principled transfer of information and uncertainty between sub-models targeting complementary aspects of a common process or system, as one of the statistical legacies of the pandemic. It will allow tractable models of complex systems to be developed in a modular way and will have wide applicability for tackling future challenges facing science and society, such as climate change and consequent economic and behavioural modelling.

6 | THE CHANGING WORLD OF DATA-DRIVEN ACTIVITIES

Our world increasingly relies on data driven activities, the management of the Covid-19 crisis being a recent case in point. This has led to a shift and questioning of the place of statistics as a discipline in such activities. In his seminal '50 Years of data science' paper, Donoho (2017) discusses the multiple facets of data-driven activities that are at the core of data science and concludes by proposing 'the science of learning from data' as a formulation that captures their essence, a proposition that I adhere to. In 2018, the US National Foundation for Sciences sponsored a workshop 'Statistics at a crossroads: who is for the challenge?', which brought together leading researchers and educators in statistics tasked with developing a 10- to 20-year vision for the field of statistics. The executive summary of the workshop report (NSF, 2019) included the stark warning: 'we either flourish by embracing and leading data science or we decline and become irrelevant. We must evolve and grow to be the transdisciplinary science that collects and extracts useful information from data'. Embracing data science is the overall vision behind the creation of the Harvard Data Science Review. In the editorial for its inaugural volume, the Editor-in-Chief, Xiao-Li Meng argues that 'data science' has emerged as an umbrella term to capture a large and evolving (artificial) ecosystem of data driven activities, mix of data engineering and wrangling, statistics (modelling and inference), machine learning, computing and visualisation, strongly focussed on practice and impact, on multidisciplinarity and close interactions with domain fields (Meng, 2019). The success of the Harvard Data Science Review with its broad-church outlook is a testimony to his vision.

It will come as no surprise that the changing data landscape in which statisticians operate has been a recurrent theme of previous presidential addresses. David Hand describes the breadth of modern statistics and its roots, emphasising the dramatic impact of computers on the evolution of our discipline, and refers to the inclusivity of 'greater statistics', a term coined by Chambers (1993) to mean 'everything related to learning from data' as the outward looking direction for statistics (Hand, 2009). Peter Diggle firmly argues for embracing data science and for proudly asserting what we can contribute to it, stressing the fundamental need to use probability to deal with uncertainty, the ubiquitous relevance of statistics to the whole of natural and social science and our role of statistical scientists embedded into a range of scientific disciplines (Diggle, 2015). Taking a historical perspective, Deborah Ashby stresses the continuity of our mission in exploiting data for the public good and champions actions to grow our capacity and our influence in response to the explosion in data (Ashby, 2019).

It is thus in line with my eminent predecessors that I set in motion an RSS Data Science Task Force at the beginning of 2021 to devise how the RSS can develop an overall strategy of visible engagement with data science, which builds on the strong growth in this area and on the existing work of Sections and Special Interest Groups, which have been active in this field for some time to the Society's great benefit. After a year of brainstorming, the Task Force recommended three major areas of investments to upgrade our offering to our members as well as to position the RSS as a leader in the wider field of data science. The priorities agreed by the group include: (i) fostering a highly skilled, diverse data science workforce that meets the future needs of our economy and society; (ii) offering data science practitioners additional online resources to increase the impact of data science applications; and (iii) promoting cutting edge pan-data-science methodology. The Task Force judged that the RSS is well positioned to achieve these aims as it can build on the diversity of its membership, which include researchers working at the cutting edge of statistical science, data science practitioners in leading roles in data-intensive businesses, and leaders of data science in the public sector, and on its strong track record in academic publishing, industry networking and professional development and accreditation.

After a little more than a year since the launch of this strategic initiative, I am gratified to report that following council approval in March 2022, concrete steps to implement all three identified priorities are underway. Priority (i) is the most advanced, embodied by the *Alliance for Data Science Professionals*, a network of professional organisations[4] and employers united in the goal to develop a joint framework for cross-organisation accreditation and standards for data scientists, covering both skills and professional conduct. This initiative, started with forethought in early 2020, has been led by the RSS VP of Professional Affairs and the Head of Professional Affairs. As I am writing, professional standards are already agreed by Alliance members and beta testing of the certification process is taking place with the aim of formally starting accreditation in the autumn 2022. Through the diversity of the member societies comprising the Alliance, this initiative exemplifies the broad reach of data science and the benefits of fostering a collaborative professional ecosystem.

Priorities (ii) and (iii) are currently spearheaded by the active engagement of RSS fellows. With support from the Turing Institute, the data science section launched a survey of data scientists, to guide the content of a new RSS 'Data Science Central' platform designed to contain resources of practical benefit to data science practitioners, such as case studies. This central platform will foster dynamic interactions between data scientists, machine learners and the RSS so that we can keep tailoring the relevance of our offering and help build communities of good practice.

Finally, to fulfil priority (iii), a major initiative, the creation by the RSS of a new online open access data science journal, was supported by Council. The vision for this new journal is to span the data science fields and provide a natural forum for new approaches and findings that would particularly benefit from reaching a broad audience. Hot topics, such as, how to integrate knowledge in the wide sense into deep learning machinery, how to ensure not only reliability and trustworthiness of algorithms but also encode ethical constraints in learning schemes, or how to derive simultaneously the optimal data acquisition and estimation sequence to achieve specific real-world goals, come to mind among many other topics that would flourish from the existence of such a publication outlet. Currently the excellent journals and conferences serving data science are mostly specific to a sub-field (mathematical statistics, machine learning, econometrics, etc.) or an application domain (computer vision, biology, etc.). This leads to fragmentation of the data science field and difficulties for statisticians to embrace and interact creatively with recent developments. As clearly conveyed by Peter Green in his address, diversity is to be celebrated but fragmentation needs to be combatted by strengthening the bonds between different parts of the discipline (Green, 2003). By creating a new data science journal with a pan-data science ambition, the RSS will emphasise cross-fertilisation and unifying themes in how we learn from data, taking the lead in filling an important gap in the diffusion of innovative ideas and tools to a diverse community.

There will be no doubt challenges to overcome to ensure the long-term success of our data science investment, but I believe that our three-pronged strategy is timely and responsive. Continuing to help transform this vision into reality is high on my agenda and I am extremely grateful for the time and energy that data science colleagues have devoted to this task over the last months. Their enthusiastic commitment has given the RSS important strategic keys to go forward that are commensurate with the breadth and depth of developments in the data science field.

7 | CONCLUDING REMARKS

As I write, we appear to have entered a new phase in the management of the Covid-19 pandemic. Many uncertainties remain on the long-term consequences of the pandemic and solid evidence will continue to be needed in looking back in terms of lessons learnt and forward in terms of future preparedness. My personal engagement in the statistical issues of the pandemic, which arose both naturally as a continuation of my lifelong statistical work in health, and as a necessity commensurate with our societal responsibility as a professional society, has informed the reflections that I have shared. Despite the deliberately chosen angle on Covid-19, the lessons that we can draw on the principles of engagement, the role of statisticians faced with a societal emergency, and the means to increase our capacity to impact are general and can help us prepare for what the future will throw at us. We are entering a period of major world uncertainties in the health, environment, economic and political arenas. Some key areas such as climate change are already at the forefront, but others might emerge in a short time scale. The RSS has recently put into place a Campaign Advisory Group which will navigate the dual goals of being responsive to important opportunities for influencing public policies or public debate as well as focus on a few key strategic areas such as Equality and Diversity and Statistical Literacy. In parallel, I believe that a focus of our community on statistical agility in its broadest sense will become increasingly needed to tackle the substantial uncertainties facing us. By statistical agility, I am referring equally to theoretical advances in approximate inference, to the development of modular and interoperable frameworks and computations, to innovations in adaptive evaluation methods targeting products, systems and policies, and to many other aspects that I do not have the space to list. The pivotal strategic engagement of the RSS towards data science that I discussed earlier will be key to advance the many facets of statistical agility.

It is with great sadness and a deep sense of shared loss[5] that we miss Sir David Cox's insightful voice to remind us of what is fundamentally important for our mission in today's uncertain world. I will end with a quote from his 1981 presidential address: 'In principle, decisive actions can be combined with intellectual appreciations that there are uncertainties in the key evidence in which decision making is based. In theoretical terms these are the dual themes of decision and inference that reappear throughout recent discussions of general principles in statistics. The possibility of explicit and quantitative resolution of this conflict is one of the most important intellectual contributions to our subject, with far-reaching and as yet undeveloped applications' (Cox, 1981). His prescient thoughts and encouragements to pursue with enthusiasm our vision of using data and our unique skills for the benefit of science and society will resonate for many years to come.

ACKNOWLEDGEMENTS

I am extremely grateful to Arnoldo Frigessi, Peter Green and Stian Westlake for their encouragements and helpful comments on an early draft of this address. A special thanks to all the colleagues with whom I have worked closely on Covid-19-related issues, in particular colleagues in the RSS Covid-19 Task Force, in the MRC Biostatistics Unit and in the Turing-RSS Health Data Lab. The stimulating discussions that we have had over the last 2 years have been crucial in shaping these thoughts. The responsibility for the views expressed is entirely mine.

ENDNOTES

[1] The so-called 3Vs of big data introduced in 2001 by Doug Laney.

[2] People in the community or in the health care system who have had at least one positive Covid-19 test result.

[3]This project was part of a broad partnership between the Alan Turing Institute, the Royal Statistical Society and the Joint Biosecurity Centre (JBC), to provide additional capacity to JBC and subsequently UKHSA through open-science research based on rigorous statistical modelling and inference. See Sections 3.3 and 4.3 for further details on the Health Data Lab created through this partnership.
[4]British Computing Society, Institute for Mathematics and its applications, National Physical Laboratory, Operational Research Society, the Alan Turing Institute, Royal Statistical Society.
[5]Sir David Cox died on the 18 January 2022.

REFERENCES

Ashby, D. (2019) Pigeonholes and mustard seeds: growing capacity to use data for society. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182(4), 1121–1137.

Aylin, P., Best, N., Bottle, A. & Marshall, C. (2003) Following Shipman: a pilot system for monitoring mortality rates in primary care. *The Lancet*, 362(9382), 485–491.

Bird, S.M., Cox, S.D., Farewell, V.T., Harvey, G., Tim, H. & Smith, P.C. (2005) Performance indicators: good, bad, and ugly. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1), 1–27.

Birrell, P., Blake, J., Van Leeuwen, E., Gent, N. & De Angelis, D. (2021) Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. *Philosophical Transactions of the Royal Society B*, 376(1829), 20200279.

Blastland, M., Freeman, A.L., van der Linden, S., Marteau, T.M. & Spiegelhalter, D.J. (2020) Five rules for evidence communication. *Nature*, 587, 362–364.

Brookmeyer, R. (2010) Measuring the HIV/AIDS epidemic: approaches and challenges. *Epidemiologic Reviews*, 32(1), 26–37.

Brookmeyer, R. (2021) Commentary on the role of statisticians in pandemics. *Statistics in Medicine*, 40(11), 2521.

Brookmeyer, R. & Gail, M.H. (1988) A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *Journal of the American Statistical Association*, 83(402), 301–308.

Carmona, C. & Nicholls, G. (2020) Semi-modular inference: enhanced learning in multi-modular models by tempering the influence of components. *Proceedings of the international conference on artificial intelligence and statistics*. PMLR, pp. 4226-4235.

Chambers, J.M. (1993) Greater or lesser statistics: a choice for future research. *Statistics and Computing*, 3(4), 182–184.

Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K. et al. (2020) Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*, 25(3), 2000045.

Covid 'Lessons learned memo'. (2021) *Ten statistical lessons the government can learn from the past year*. Available at: https://rss.org.uk/RSS/media/File-library/Policy/2021/RSS-Stats-Data-and-Covid-FINAL-embargoed-09-03-21.pdf

Cox, D.R. (1981) Theory and general principle in statistics. *Journal of the Royal Statistical Society: Series A (General)*, 144(3), 289–297.

Curnow, R.N. (1999) Unfathomable nature and government policy. *Journal of the Royal Statistical Society Series D (The Statistician)*, 48(4), 463–476.

De Gruttola, V. & Lagakos, S.W. (1989) Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, 45, 1–11.

Di Domenico, L., Pullano, G., Sabbatini, C.E., Boëlle, P.Y. & Colizza, V. (2021) Modelling safe protocols for reopening schools during the COVID-19 pandemic in France. *Nature Communications*, 12(1), 1–10.

Diggle, P.J. (2015) Statistics: a data science for the 21st century. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(4), 793–813.

Donoho, D. (2017) 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), 745–766.

Ellenberg, S.S. & Morris, J.S. (2021) AIDS and COVID: a tale of two pandemics and the role of statisticians. *Statistics in Medicine*, 40(11), 2499–2510.

Gallo, R.C. & Montagnier, L. (1988) AIDS in 1988. *Scientific American*, 259(4), 40–51.

Goldstein, H. & Spiegelhalter, D.J. (1996) League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159(3), 385–409.

Gore, S.M. & Armitage, P. (1988) Royal statistical society meeting on AIDS. *Journal of the Royal Statistical Society Series A*, 151(1), 3–4.

Goubar, A., Ades, A.E., De Angelis, D., McGarrigle, C.A., Mercer, C.H., Tookey, P.A. et al. (2008) Estimates of human immunodeficiency virus prevalence and proportion diagnosed based on Bayesian multiparameter synthesis of surveillance data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3), 541–580.

Goudie, R.J., Presanis, A.M., Lunn, D., De Angelis, D. & Wernisch, L. (2019) Joining and splitting models with Markov melding. *Bayesian Analysis*, 14(1), 81.

Green, P.J. (2003) Diversities of gifts, but the same spirit. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(4), 423–438.

Halloran, M.E. (2021) Comment on AIDS and COVID-19: a tale of two pandemics and the role of statisticians. *Statistics in Medicine*, 40(11), 2524.

Hand, D.J. (2009) Modern statistics: the myth and the magic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(2), 287–306.

Hernán, M.A. & Robins, J.M. (2020) *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC.

Holt, D.T. (2008) Official statistics, public policy and public trust. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2), 323–346.

Jewell, N.P. (1994) Non-parametric estimation and doubly-censored data: General ideas and applications to AIDS. *Statistics in Medicine*, 13, 2081–2095.

Johns Hopkins University Coronavirus Resource Center (n.d.) Available at: https://coronavirus.jhu.edu/

Kingman, J. (1989) Statistical responsibility. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 152(3), 277–285.

Kingman, J., Durbin, J., Cox, D. & Healy, M.J.R. (1988) Statistical requirements of the AIDS epidemic. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 151(1), 127–130.

Kunzmann, K., Lingjaerde, C., Bird, S. & Richardson, S. (2022) The how matters: simulation-based assessment of the potential contributions of lateral flow device tests for keeping schools open and COVID-safe in England. *Harvard Data Science Review*, 4(1). https://doi.org/10.1162/99608f92.5a0c7aaf

Lawrence, N. & Montgomery, J. (2020) *Storming the castle: data science for Covid-19 policy*. https://www.bennettinstitute.cam.ac.uk/blog/storming-castle-data-science-covid-19-policy/

Lee, L.Y., Rozmanowski, S., Pang, M., Charlett, A., Anderson, C., Hughes, G.J. et al. (2022) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infectivity by viral load, S gene variants and demographic factors, and the utility of lateral flow devices to prevent transmission. *Clinical Infectious Diseases*, 74(3), 407–415.

Leng, T., Hill, E.M., Holmes, A., Southall, E., Thompson, R.N., Tildesley, M.J. et al. (2022) Quantifying pupil-to-pupil sars-cov-2 transmission and the impact of lateral flow testing in English secondary schools. *Nature Communications*, 13(1), 1–11.

Medical Research Council. (1948) Streptomycin treatment of pulmonary tuberculosis: a medical research council investigation. *BMJ*, 2, 769–782.

Meng, X.L. (2019) Data science: an artificial ecosystem. *Harvard Data Science Review*, 1, 1.

Moser, C. (1980) Statistics and public policy. *Journal of the Royal Statistical Society: Series A (General)*, 143(1), 1–32.

Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K.B., Jersakova, R., Lomax, J. et al. (2022a) Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework. *Nature Microbiology*, 7(1), 97–107.

Nicholson, G., Blangiardo, M., Briers, M., Diggle, P.J., Fjelde, T.E., Ge, H. et al. (2022b) Interoperability of statistical models in pandemic preparedness: principles and reality. *Statistical Science*, 37(2), 183–206.

Nogrady, B. (2021) 'I hope you die': how the COVID pandemic unleashed attacks on scientists. *Nature*, 598(7880), 250–253.

NSF Report. (2019) *Statistics at a crossroads: who is for the challenge?* Available at: https://www.nsf.gov/mps/dms/documents/Statistics_at_a_Crossroads_Workshop_Report_2019.pdf

Nyberg, T., Twohig, K.A., Harris, R.J., Seaman, S.R., Flannagan, J., Allen, H. et al. (2021) Risk of hospital admission for patients with SARS-CoV-2 variant B. 1.1. 7: cohort analysis. *British Medical Journal*, 373, n1412.

Nyberg, T., Ferguson, N.M., Nash, S.G., Webster, H.H., Flaxman, S., Andrews, N. et al. (2022) Comparative analysis of the risks of hospitalisation and death associated with SARS-CoV-2 omicron (B. 1.1. 529) and delta (B. 1.617. 2) variants in England: a cohort study. *The Lancet*, 399(10332), 1303–1312.

Our World in Data. (2022) *Our world in data coronavirus pandemic (Covid-19) dashboard*. Available at: https://ourworldindata.org/coronavirus#explore-the-global-situation

Pellis, L., Birrell, P.J., Blake, J., Overton, C.E., Scarabel, F., Stage, H.B. et al. (2022) Estimation of reproduction numbers in real time: conceptual and statistical challenges. *Journal of the Royal Statistical Society: Series A* to appear.

Poole, D. & Raftery, A.E. (2000) Inference for deterministic simulation models: the Bayesian melding approach. *Journal of the American Statistical Association*, 95(452), 1244–1255.

Presanis, A.M., Pebody, R.G., Birrell, P.J., Tom, B.D., Green, H.K., Durnall, H. et al. (2014) Synthesising evidence to estimate pandemic (2009) A/H1N1 influenza severity in 2009–2011. *Annals of Applied Statistics*, 8(4), 2378–2403.

Pullinger, J. (2013) Statistics making an impact. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(4), 819–840.

Rossman, H. & Segal, E. (2022) Nowcasting the spread of SARS-CoV-2. *Nature Microbiology*, 7(1), 16–17.

Royal Society DELVE Initiative. (2020) *Balancing the risk of pupils returning to schools*. Available at: https://rs-delve.github.io/reports/2020/07/24/balancing-the-risk-of-pupils-returning-to-schools.html

rss.org. (2021) *Royal statistical society diagnostic tests working group report*. Available at: https://rss.org.uk/RSS/media/File-library/Policy/2021/RSS-Diagnostic-tests-report-FINAL.pdf

Shi, X., Pan, Z. & Miao, W. (2022) Data integration in causal inference. *WIREs Computational Statistics*, e1581. https://doi.org/10.1002/wics.1581

Singanayagam, A., Hakki, S., Dunning, J., Madon, K.J., Crone, M.A., Koycheva, A. et al. (2022) Community transmission and viral load kinetics of the SARS-CoV-2 delta (B. 1.617. 2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study. *The Lancet Infectious Diseases*, 22(2), 183–195.

Smith, A.F. (1996) Mad cows and ecstasy: chance and choice in an evidence-based society. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159, 367–384.

Spiegelhalter, D.J. (2017) Trust in numbers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(4), 948–965.

Spiegelhalter, D.J. & Best, N. (2004) Shipman's statistical legacy. *Significance*, 1(1), 10–12.

Spiegelhalter, D.J. & Masters, A. (2021) *Covid by numbers: making sense of the pandemic with data*. Penguin Random House.

Spiegelhalter, D.J., Aylin, P., Best, N.G., Evans, S.J. & Murray, G.D. (2002) Commissioned analysis of surgical performance using routine data: lessons from the Bristol inquiry. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2), 191–221.

Young, B.C., Eyre, D.W., Kendrick, S., White, C., Smith, S., Beveridge, G. et al. (2021) Daily testing for contacts of individuals with SARS-CoV-2 infection and attendance and SARS-CoV-2 transmission in English secondary schools and colleges: an open-label, cluster-randomised trial. *The Lancet*, 398(10307), 1217–1229.

Yu, B. & Singh, C. (2022) Seven principles for rapid-response data science: lessons learned from Covid-19 forecasting. *Statistical Science*, 37(2), 266–269.

How to cite this article: Richardson, S. (2022) Statistics in times of increasing uncertainty. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 1–25. Available from: https://doi.org/10.1111/rssa.12957