SIR DAVID R. COX

15 July 1924 — 18 January 2022

D.R. Cox

SIR DAVID R. COX

15 July 1924 — 18 January 2022

Elected FRS 1973

BY NANCY REID FRS*,†

Department of Statistical Sciences, University of Toronto, 9th Floor, 700 University Avenue, Toronto, Canada M5G 1Z5

David Cox was the leading statistical scientist of his generation and had an extraordinary influence on the field. His research career spanned some 75 years and 386 published works, with several drafts in preparation at the time of his death. He was held in extremely high regard by the community of scholars for his seminal contributions to scholarship, his enthusiasm for science and his generosity of intellect. One obituary called him a 'rock star' of the statistical world, and in spite of the hyperbole the description is apt. His work was very broad; he made influential contributions to the fields of experimental design, stochastic processes, statistical methodology, foundations of inference, statistics in medicine and public health, and more. His 23 published books continue to be key references for students and researchers. His most widely cited paper (*J. R. Stat. Soc. B* **34**, 187–220 (1972)) introduced what is now called the Cox model for the analysis of survival data; this was included in *Nature*'s list of the top 100 cited scientific papers of all time. He received many accolades, including the Copley Medal (2010) and the inaugural International Prize in Statistics (2016).

EARLY LIFE

David Roxbee Cox was born in Birmingham, the only child of Sam Roxbee Cox and Lilian Esther Cox (née Braines). His father was in the family jewellery business located in Birmingham's famous jewellery quarter. His mother was a homemaker, and took a keen

*Email: nancym.reid@utoronto.ca

†Sections of this memoir closely follow text in Davison *et al.* (2022) and Battey & Reid (2022), and I thank my collaborators for their generosity.

interest in David's education. David attended St Michael's Church School in Handsworth (1929–1935), and Handsworth Grammar School (1935–1942), the latter on a scholarship. He remarked that he learned nearly nothing during his years at St Michael's, except through his mother's tuition. He spoke very highly of the teaching at Handsworth Grammar, especially in mathematics, physics and chemistry. His autobiographical summary at the Royal Society mentions that the school was evacuated to Stroud and then Barnstable in 1939; this must have made a very large impression. I once fretted in his presence about the state of the world, and he replied with some indignation, 'I don't see any bombs dropping'. David was one of very few boys from Handsworth to go to university; he left school with a State Scholarship, a City Major Scholarship and an Open Major Scholarship in natural sciences at St John's College, Cambridge. He said that education was considered very important in his home; he thought his mother would have very much liked to be a school teacher. She collected, and saved, various pieces about his school career, which indicate that he was an excellent student in nearly every subject.

University education

At Cambridge David took a first class in the Preliminary Examinations and in Part II of the Mathematical Tripos, and attended the Part III lectures. He was inspired by a course in applied mathematics from Paul Dirac FRS; David's interests at the time were in pure mathematics and mathematical physics. He had lectures in statistics from Harold Jeffreys FRS and J. O. Irwin. He commented that the undergraduate lectures on statistics at Cambridge were relatively uninspiring (57)[*], but he often enjoyed recalling professors there who, he claimed, could write on the blackboard with one hand while erasing parts with the other and at the same time explaining a third topic.

In 1944, following two years of his undergraduate degree, he was seconded to wartime service at the Royal Aircraft Establishment (RAE), working on strength of materials in a department of structural and mechanical engineering. As the war was coming to an end, he happened to see an advertisement for a research position at the Wool Industries Research Association (WIRA) under the supervision of Henry Daniels (FRS 1980). He had been studying a paper by Daniels as part of his work at the RAE, so, instead of returning to Cambridge as he was expecting, he took this position in Leeds, which started him on a lifetime of research in stochastic processes and statistical science.

David completed his PhD in 1949 at the University of Leeds, with B. L. Welch as nominal supervisor, although his thesis was based on the work at WIRA under Daniels' supervision. With the encouragement of Daniels he attended lectures by Maurice Bartlett (FRS 1961) in Manchester; he described the trip from Leeds to Manchester at that time as 'a major journey, like going to the North Pole' (Reid 1994).

Leadership and career

In his Royal Society autobiographical summary, David lists his position at WIRA (1946–1950) as junior, then senior scientific officer and, after his first six months, head of mathematics and

[*] Numbers in this form refer to the bibliography at the end of the text.

Figure 1. The staff of the Cambridge Statistical Laboratory, 1954. David is in the front row, fourth from left. The complete list of staff is available at Cambridge Statistical Laboratory (2019). (Courtesy of Statistical Laboratory, Cambridge.)

statistics. From 1950–1955 he was an assistant lecturer in the Statistical Laboratory at the University of Cambridge (figure 1). When this position ended there was no permanent position available in Cambridge, and David and his wife and children spent a year in the United States. In 1956 he was appointed reader at Birkbeck College, London, and in 1960 became professor and head of statistics. From 1966 to 1988 he held the chair in statistics at Imperial College, London, and in 1988 he moved to Oxford to become warden of Nuffield College. Although he officially retired in 1994, he continued with research, supervision and teaching at Nuffield and in the Oxford Department of Statistics. He commuted daily to his office at Nuffield until the disruption of the pandemic in the spring of 2020.

That he was appointed head of his research group while in his twenties is not in retrospect surprising; he was a natural academic leader. He had good judgement and was quick and decisive when addressing problems of an administrative nature. His manner was inimitably understated, yet he had an uncanny ability to delegate in a way that ensured the delegates carried out their assigned tasks more or less exactly as he wished: 'Don't you agree it would be a good idea to . . . ?' David's colleague Professor John Goldthorpe has remarked that David's wardenship will be seen as marking a crucial turning point in the history of the college. He also noted that David dispatched the day's administrative tasks between 08.00 and 09.00, leaving the remainder of the day for research and discussion with colleagues.

While at Imperial College, David served as head of the statistics group in the Department of Mathematics, and as head of department (1969–1973). The statistics group at Imperial had a high proportion of women, especially unusual for the time. David also had a relatively high proportion of women among his collaborators; he took women seriously, which was a refreshing attitude in academia at that time. A special issue of the *Harvard Data Science Review* (Richardson & Wermuth 2023) includes many reminiscences from postgraduate students at Imperial College, and these essays give a flavour of his impact on the research group. Regular attendance at Wednesday presentations at the Royal Statistical Society was essentially mandatory—the students hurrying down the street behind David to the London Underground.

David's international reputation was such that visitors from around the world regularly planned their trips through London, to visit the department for an hour, a day or several months. After he moved to Oxford, he welcomed a similarly steady stream of visitors. The many personal tributes circulating after his death invariably mentioned his warm welcome and exceptional generosity with his time and ideas.

In addition to the leadership he provided at his home institutions, he engaged in a great deal of professional service. He was president of the Bernoulli Society, the Royal Statistical Society (RSS) and the International Statistical Institute. As president of the RSS he oversaw the preparations for the 150th anniversary of the society in 1984. He was editor of the *Journal of the Royal Statistical Society* from 1960 to 1964.

David is most closely associated with *Biometrika*, which he edited from 1966 to 1991, famously reading every paper and often sending helpful comments to the authors, whether or not their papers were accepted. Under his editorship, *Biometrika* became the leading venue for new work in statistical theory and methods. He continued to provide leadership to the journal by serving as chair, and then member of the Biometrika Trust (1992–2019).

Family life

David married Joyce Drummond (b. 1924) in Keighley, Yorkshire, in 1948. They met through their work in the wool industry; an early publication (2) acknowledges the valuable help of Mrs J. Cox. Joyce was the only child of John and Edith Drummond; John was a buyer in the wool trade and was originally from Scotland. David and Joyce had four children: Joan (1949), John (1953), Andrew (1959) and Steven (1961) (figure 2).

Following their year in the USA, David and Joyce settled in the London suburb of Eastcote in 1956, where they stayed until their move to Oxford in 1988. In Eastcote they provided a stable environment for their children, based on firm values such as the commitment to truth and a certain frugality. His children recall with humour their parents' extreme reluctance to invest in a new television, even when the old set gave a barely watchable picture and emitted a loud hum. Always generous, thoughtful and constant, his children remember him as a kind of lodestar in a world in flux. He grew older but in a way seemed unchanged and his interest in ideas never seemed to wane.

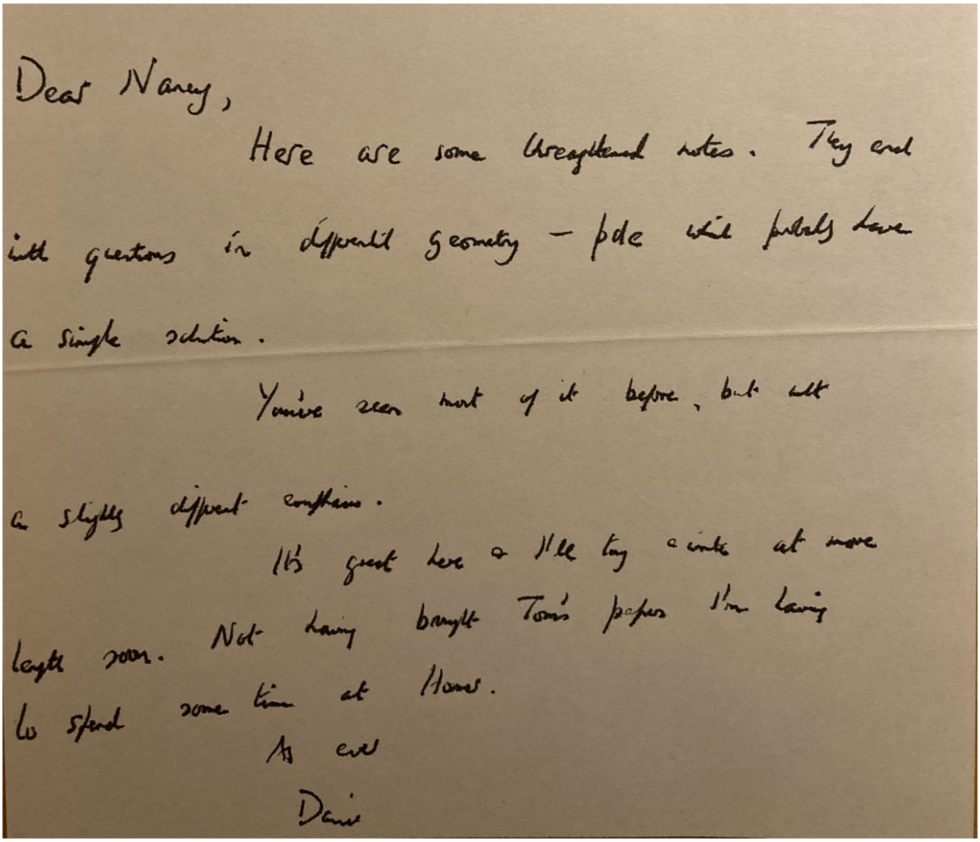
David's belief in the scientific method informed his life: he was committed to rationalism, sceptical of religion and internationalist in outlook. His political views reflected those of the left wing in post-war Britain. He took his work very seriously but often appeared to be mildly amused at much of human activity. The absurdity of language in the media often made him laugh, and he loved a well-planned pun. Self-effacing and modest about his own achievements, he seemed slightly bemused by the awards that were showered upon him in later life. He preferred his study to the limelight.

Figure 2. David, Joyce and family on the occasion of David's ninetieth birthday.

David was very widely read; a literary favourite he often mentioned was George Eliot's *Middlemarch*, but his taste ran more to non-fiction (figure 3). His children recall growing up in a house full of books, dictionaries and reference books on all sorts of subjects. He could converse in depth on subjects ranging from politics, geography, history, observations from travels, early cinema and jazz, to the state of the Aston Villa football team and train-spotting. His interest in psychology and the unconscious is described in Battey (2023). He had broad interests in classical music, especially the operas of Mozart and Wagner (Isham 2023).

In statistical matters 'his opinions were strong, but mildly expressed' (Davison 2023). In matters of politics, in my experience his opinions were strong and forcefully expressed, but never overbearing. His life-long support of Labour reflected and drove a very egalitarian approach to his interactions in statistics, described in Kartsonaki (2023): 'he valued other people's time and often thought that people who were just starting their careers had valuable contributions to make.'

He drew the most satisfaction from working, and his enthusiasm for this was infectious. One felt in his presence that the statistical and scientific ideas under discussion were more interesting and important than anything else. In a video interview, when asked about his hobbies, he replied: 'Hobbies? There's no time for hobbies! [. . .] I do a certain amount of non-scientific reading, but I wouldn't quite call that a hobby [. . .] Life's too short!' (Valsecchi 2014).



Dear Nancy,

Here are some [illegible] notes. They end with questions in differential geometry – pde which probably have a simple solution.

You've seen most of it before, but with a slightly different emphasis.

It's great here & I'll try & write at more length soon. Not having brought Tom's papers I'm having to spend some time at Homer.

As ever

David

Figure 3. Until about 1990 academic exchanges were all by written letter. This note was sent from Australia in January 1987. I neglected to enquire as to whether David managed to finish Homer.

Research activities and achievements

Introduction

Bradley Efron, who shared the BBVA Foundation Prize with David Cox in 2016, wrote (Efron 2023):

> It was said that Mozart could write all 17 parts of symphony at once, [. . .] David Cox's synoptic view of statistics was Mozartian in the same sense. Statistics is a messy field, with its working parts scattered across the scientific world, but David seemed to be able to take it in all of a piece.

Statistical science takes its *raison d'être* from its applications to nearly every field of scientific enquiry. New developments in the theory of statistics are measured against their utility, and the best theoretical work is developed to unify a set of specific applied problems. In practice, though, most statistical science researchers tend to work on either the theoretical or the applied side of the discipline—there is not time to do everything. David was one of a few, and of these few the best, who moved with ease between theory and applications. About one of his most widely-known contributions to the theory of inference, he said (Reid 1994):

> it was strongly motivated by practical experience, and yet on the other hand I couldn't say it arose from one particular type of applied problem. It arose in a sense from all the applied work I'd done to that point.

David's research programme included a large body of work in the field of stochastic processes, following in the long tradition of applied mathematics at Cambridge and influenced by the pioneering work of Maurice Bartlett and Henry Daniels. While modern research in stochastic analysis and statistical science has very little overlap, this was not the case in David's 'synoptic view'.

The following sections follow David's research in roughly chronological order, although the boundaries between subjects and over time are inevitably blurred, not least because David's interest was in advancing science with however many 'working parts' might be needed.

Wool

The WIRA was established in 1918; in 1989 it merged with the Shirley Institute to become the British Textile Technology Group. It provided testing and certification for the wool industry, and had a prominent and successful research programme. Two chemists, A. J. P. Martin (FRS 1950) and R. L. M. Synge (FRS 1950), were awarded the Nobel Prize in Chemistry for their development of partition chromatography. According to David (41), Henry Daniels made key contributions to this research. WIRA seems to have been an institution for which there is no modern equivalent. David said the method of the director of research, Mr H. B. Wilsdon, was 'to appoint good people, to give them some broad encouragement, and then to give them much freedom to pursue their own ideas'.

David's experience at WIRA profoundly influenced his research throughout his life. 'In wool research, you see, you've got everything: from the biology and nutrition of the sheep, through the chemistry and physics of various processes, to the operational research side, the engineering side, and the economic side. So there was an enormous range of problems and extremely good people working there' (Reid 1994). Between 1947 and 1952 he published a book, 15 papers and four government reports. The short book, with A. Brearley (1), went to a fifth edition in 1960. David's first read paper to the RSS (3) acknowledges stimulating discussion with Mr H. B. Wilsdon.

Most of David's early papers were motivated by problems encountered at WIRA, and address design of experiments, sequential sampling and, by 1954, various types of stochastic processes.

Stochastic processes

The RSS has a long tradition, which continues, of evening presentations based on papers that have been circulated as preprints. After the presentation, two reviewers propose, and then second, a vote of thanks. This is followed by short oral contributions to the discussion, and all of this is published with the paper, along with received written contributions to the discussion and a reply from the author(s). In the time of David's early career it was often the case that the ostensibly polite proposer and seconder were highly critical of the paper and its presentation, in a sometimes acerbic style that readers outside the UK might think of as 'typically British'.

In 1955 David delivered his first read paper (3) on 'Some statistical methods connected with series of events'. This paper introduced what doubly stochastic Poisson processes, or Cox processes. The homogeneous Poisson process is the basic model for complete randomness of events observed in time, or space, or both. Such a process (in time) is described by the conditions that (i) the instantaneous rate of the process is a positive constant λ and (ii) events in non-overlapping intervals occur independently. A non-homogeneous Poisson process allows the rate of the process to vary with time, $\lambda = \lambda(t)$, and/or with related measured variables, $\lambda = \lambda(t;\ x)$. Examples of applications discussed include electrical pulses in a nerve fibre, customers arriving in a queue, emissions of radioactive particles and *slubs* (imperfections) in a length of wool yarn (3). The doubly stochastic Poisson process, now called the Cox process, models the rate function itself as a random process. Modern areas of application of the Cox process include hydrology, geology, finance, epidemiology and machine learning. Versions in which the logarithm of the rate function is a Gaussian process, log-Gaussian Cox processes, are now routinely used in applications thanks to computational advances that were unthinkable when the paper was written.

Other papers by David in the 1950s on stochastic processes included three papers on renewal theory and the superposition of point processes, written with Walter Smith, a Cambridge PhD student supervised jointly by David and Henry Daniels. A few years later, a monograph by Cox and Smith on queueing theory was published (7). The pattern of forming a collaboration with an ex-research student, culminating in the publication of a research monograph, was to be repeated many times during David's career. Just a year later, David published a second monograph on renewal theory (9), and followed this not long afterwards with a third on the statistical analysis of series of events (11).

During the first half of the 1960s, David also wrote a ground-breaking textbook on stochastic processes (10) with Hilton Miller, a colleague of David's at Birkbeck and later at Imperial College. The text, intended for mathematics and statistics students and for research workers, introduces the main mathematical techniques useful for analysing stochastic processes that arise in a broad range of scientific applications. Thus, in the five-year period between 1961 and 1966 and in addition to 23 (mostly single-author) papers on a range of other topics, David published three monographs encapsulating the results of previous research in a highly succinct and accessible form, together with a substantial textbook on stochastic processes. These books have become literature standards, used by generations of postgraduate students.

A long and fruitful period of collaboration and friendship followed a serendipitous meeting in 1986 with Ignacio Rodríguez-Iturbe, a distinguished Venezuelan hydrologist (Levin & Rinaldo 2022). At a banquet to celebrate David's eightieth birthday, Rodríguez-Iturbe described the accidental meeting: 'are you *the* David Cox?'—the reply reportedly, 'guilty as charged'—and his delight at the prospect of working with the great man. Valerie Isham, David and Ignacio had a remarkable and productive 20-year collaboration in statistical hydrology, described in Davison *et al.* (2022) and Isham (2023) (figure 4).

Their series of papers first extended a highly idealized model for the spatial distribution of total storm rainfall, to more complex spatio-temporal processes, first at a single site (22, 25), and later to cluster processes incorporating both space and time (24, 33, 35). The important feature of these models is that the individual rainfall cells, with locations governed by a spatial point process, also have random but finite temporal durations and spatial extents, enabling realistic representation of dry periods and/or regions. These models have helped to solve a range of catchment-based hydrological problems linking stochastic models for rainfall fields to distributed models for soil run-off and stream flow (36, 45), have been successfully applied

Figure 4. Valerie Isham, David Cox and Ignacio Rodríguez-Iturbe in Caracas in 1987. (Courtesy of Valerie Isham.)

in many different climatological regimes and have had a substantial impact on hydrological practice and policy.

In hydrology, as in much of his applied work, David's characteristic approach was analytical rather than numerical: first to abstract the essence of the problem, using a simple mathematical formulation of the underlying process, then to find clever and original uses of a wide range of techniques from applied mathematics to determine properties algebraically, and finally to show how these could lead to solutions of important practical problems.

David's contributions to the more classical analysis of time series include the introduction of the now-standard distinction between parameter- and observation-driven time series and a substantial discussion of data interpretation (19). Special emphasis is placed on the analysis of processes that have long-range dependence, an interest of David's since the WIRA days.

Experiments and observational studies

Design of experiments is a field of statistical science with a very long history; the terminology and main ideas were developed by Sir Ronald A. Fisher (FRS 1929) in the 1920s for analysis of agricultural field experiments at the Rothamsted Experimental Station. Randomized clinical trials are a more recent type of designed experiment, although people are much more difficult experimental subjects than are plots of a field.

David's 1958 book, *Planning of experiments* (4), is still in print and has the unusual feature among texts on statistics that it completely eschews mathematical formulas, instead expressing all the ideas through scientific common sense. His first thought had been to write up his mathematical notes from a course given in Cambridge, but he decided it would be 'much more difficult, and much more useful, to write something that was aimed at scientists with a minimum of technical statistical analysis' (Reid 1994). The Cambridge notes later formed the basis for a more theoretical treatment (37).

It was perhaps David's early practical experience at WIRA that enabled his remarkable intuition for all aspects of data collection. This influenced his writing throughout his career: for example, his ingenious approach to identifying confidence sets of models in high-dimensional regression is based on balanced incomplete block designs (58, 59) and was foreshadowed by brief comments in a much earlier paper (12). High-dimensional data problems were not unfamiliar to him—in 1962 he described computer-generated designs with more treatment factors than observations (8). His industrial experience also gave him a deep appreciation of the influence of different sources of variability. His book with Patricia Solomon (40) provides a deep but mainly non-technical account of this, illustrated with a range of applications from textile processing to microarray analysis. His book with Ruth Keogh (54) addresses the design and analysis of case–control studies. These are observational studies, usually with data collected retrospectively, in which it is desired to draw as strong conclusions as possible about the effect of exposure on disease status, while realizing the limitations entailed by not being able to randomize patients to exposure levels.

David's move to Nuffield College introduced him to many aspects of quantitative social science and thus to problems that he often remarked were very intellectually stimulating. Around this time he began an important and wide-ranging collaboration with Nanny Wermuth on multivariate data from observational studies in areas such as psychology, education and medicine, some of which is summarized in their review paper (31) and book (34). In addition to the development of graphical models and of methods for analysis of complex observational data involving both intermediate and endpoint responses, their joint work included clarification of aspects of causality (e.g. (38, 39, 42)).

Regression models

In regression analysis a response of interest is measured along with several explanatory variables in order to understand the impact on the response of variation in these variables. A response might be yield (of plants or of animals), as was common in Fisher's work, a measure of strength of samples of wool, a measure of occurrence or non-occurrence of an event, the survival time of an individual in the context of treatment for disease, and so on. The simplest model relating the explanatory variables to the response is a linear model. David's most influential methodological advances were the development of extensions of linear regression analysis to tackle a wider range of applications.

At Birkbeck College David became aware of several areas of application, particularly in psychology, that required methods of regression analysis for binary (yes/no) responses. His breakthrough then, reported in his RSS read paper (6), was to reason by analogy to least squares regression, which when adapted for binary responses leads to a logistic regression model. Like the 1955 stochastic processes paper (3), this paper is a *tour de force*. Elements of sampling theory are used to simplify the then-onerous computations involved in testing and model-fitting. The use of conditioning to focus on particular parameters of interest is derived from the exponential-family structure of the model. The paper contains much more than the idea for which it has become best known, and includes, for example, models for binary time series, as arise in psychological experiments on animal learning.

The logistic regression model is the central focus of David's 1970 monograph (13). He later commented that this book should have been written earlier, but in view of his other activities it is hard to imagine how this could have been feasible. A second edition with Joyce Snell (30) modernized the treatment to take account of the advances in computational methods then available. David's overview paper on regression analysis (12) points out that binomial, gamma and Poisson regression models are all linear exponential families, so the inferential simplicity developed for logistic regression (6) could be extended to other models.

The proportional hazards model

Data being collected in many medical applications in the 1960s and early 1970s typically involved the measurement of several patient characteristics combined with random assignment of patients to, for example, two competing treatments, and there was a pressing need for a regression model suitable for a survival time response, which is necessarily positive, but also often censored—if, for example, the patient is alive at the end of the study. David introduced a new method for regression analysis of survival data at a read paper to the RSS in 1972 (14). By this time David would have been a familiar and highly regarded figure at these meetings, and there must have been considerable excitement in advance of this particular presentation.

The first key insight was to model the hazard function, or instantaneous failure rate, instead of the distribution of the failure time; this formulation came naturally to David because of his long experience with modelling point processes. The second key insight was to introduce a flexible form for this hazard function; the multiplicative form $\lambda(t;\ x) = \lambda_0(t) \exp(x^{\mathrm{T}}\beta)$ was proposed. The baseline hazard function $\lambda_0(t)$ is common to all patients under study. The various explanatory variables x, which might be treatment, age, sex, blood pressure and so on, increase or decrease the hazard function at a rate determined by the unknown parameters β.

The stumbling block for the analysis of the data was the difficulty of estimating both the unknown function $\lambda_0(t)$, and the parameters β. The usual method would be based on the likelihood function, which is complicated and difficult to work with. The third, and critical, insight was that a novel conditioning argument could eliminate the unknown baseline function, and inference could be based on a much simpler function, later called a partial likelihood function. Proportional hazards regression is now one of the most widely used statistical methods in medicine.

David related that he worked on the paper for some years, and that the crucial insight that led to the partial likelihood analysis came to him while he was in bed with a fever and on his recovery it took some effort to recall the precise argument. Its practical value was immediately clear, and others rapidly made software available. The theoretical details of the partial likelihood construction were clarified later (16), which spurred the development of many extensions. Some of these are reviewed by Kalbfleisch & Schnaubel (2023), and others were presented at a workshop in November 2022 at the London School of Hygiene and Tropical Medicine celebrating 50 years of the Cox model.

David modestly described the 1972 paper as being cited a 'fairly large' number of times, adding the disclaimer 'although no doubt read rather less often'. It led to him being awarded the 1990 Kettering Prize and Gold Medal for Cancer Research, and stimulated major streams of research over the following decades, including the now-standard use of point process and martingale methods in the theoretical study of methods for time-to-event data and semi-parametric likelihood inference, and the application of proportional hazards and related models to data arising in medicine, public health, engineering and social science.

Statistical theory and foundations

David's 1958 paper (5) was his first to concentrate on the foundations and theory of inference. It continues to be regularly cited, as it clarifies many of the more difficult aspects of the topic.

The best-known contribution of the article is a simple example in which we suppose two measuring instruments have different precisions, the outcome of a coin toss determines which instrument is used to produce a measurement, and the measurement is used to estimate the true weight of the object measured. It is clear that the precision to be attached to the measurement is that of the instrument used; in statistical terms the appropriate inference is conditional on ancillary randomness of the coin toss. Exactly the same issue arises in many more realistic statistical models, but this example was deliberately highly stylized to illuminate the essence of the argument. The example led directly to a long and important philosophical discussion, initiated by Birnbaum (1962), on the role of the likelihood principle and the interplay between frequentist and Bayesian inference, which David revisited with the benefit of hindsight in collaboration with Deborah Mayo (50, 51). The 1958 paper also revealed that conditional inference is usually incompatible with ideas of optimality that remain popular today. The question of where to limit the conditioning is challenging; the goal is to condition on observed events to ensure that the statistical conclusions are relevant to the question at hand, while avoiding over-conditioning, which leaves too little variability for statistical conclusions (32). When the statistical model has many parameters, only some of which are of particular interest for the scientific problem, the appropriate conditional formulation becomes particularly elusive, although the conceptual argument for distinguishing samples of varying degrees of information remains compelling.

Conditional inference is just one of several topics addressed in the 1958 paper (5). David begins by distinguishing statistical inference from scientific inference, and separating the work from a decision-theoretic approach, which was very popular at the time:

> Even in problems where a clear-cut decision is the main object, it very often happens that the assessment of losses and prior information is subjective, so that it will help to get clear first the relatively objective matter of what the data say, before embarking on more controversial issues.

In a section discussing Fisher's fiducial inference, it is proposed that a set of nested confidence intervals at every confidence level provides what is called there a confidence distribution. The development emphasizes that any interval estimate, whether based on fiducial or confidence arguments, needs to have some justification in repeated sampling from the model:

> it is quite conceivable that one could construct a satisfactory measure of uncertainty that has not a direct frequency interpretation. Yet one must surely insist on some pretty clear-cut practical meaning to the measure of uncertainty

The final section addresses significance tests and *p*-values, and describes two types of null hypothesis: one distinguishing the sign of a potential effect, and a second in which the hypothesis might reasonably be expected to be exactly true. Later papers describe further types of null hypotheses. David also discusses the argument for considering the probability of an outcome ‘as or more extreme than observed’ as a measure of evidence against the null hypothesis. He continued to write on significance testing throughout his career, and published a modern and concise overview of the topic in 2020 (61). There is much ink spilled on *p*-values and significance tests in the recent scientific literature, especially associated with concerns about replicability of scientific results based on blind application of the mantra ‘*p* less than 0.05’. The clearest exposition of the issues is to be found in David’s work.

In the 1980s and 1990s there was a rapid expansion of ideas in the asymptotic theory of inference, initiated by David with O. E. Barndorff-Nielsen (17), which in turn built on Daniels (1954, 1958). David derived through formidable calculations an approximating curved exponential family, which led to what he called a local ancillary statistic (18). He also obtained an approximation to the distribution of the maximum likelihood estimator, conditional on this statistic. Several other papers in the same issue of *Biometrika* tackled related problems, and the so-called p^*-approximation emerged as a common thread. This was later linked to parameter orthogonality and approximate sufficiency for a parameter of interest in the presence of nuisance parameters (21). David’s interest in asymptotic theory was not focused on the often impressive numerical accuracy of the related approximations, but on the implications of their structure for the foundations of inference. He refused to be dazzled by intricate mathematics or clever computation, unless it was demonstrably effective for solving what he might call ‘real problems’. His pair of books with Barndorff-Nielsen (26, 32) contain a great deal of challenging mathematical detail, but are also full of statistical insight and enlightening examples.

The foundations of the theory of statistics, which David emphasized as distinct from the theoretical analysis of particular techniques, has been a difficult subject throughout the history of the discipline. One aspect of this is the ongoing discussion, which in the past was a more aggressive controversy, between Bayesian and non-Bayesian methods of inference. A sense of the controversy can be gleaned from the preface of Savage (1954):

> It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel.

Savage’s book puts forward an approach based on personalistic probability, which David reportedly described in a presentation as a ‘bold and imaginative step backwards’. His written comments on the issues are much gentler.

While his ideas on foundations did not change very much over his life, he returned to them regularly, especially in two masterful books. The text *Theoretical statistics* (15) with David Hinkley is distinguished from most books on statistical inference by its emphasis on concepts and their relevance for applications, along with a parallel de-emphasis on technical details (figure 5). It places likelihood and sufficiency at the centre of the theory, and may be the first textbook to clarify the distinction between Fisher’s significance testing and Neyman and Pearson’s accept/reject approach to hypothesis testing, treating both in considerable detail. Every potential principle of statistical inference is first explained and then challenged, so effectively that the book can seem a collection of counter-examples. This is consistent with David’s firm belief that the foundations and methods of statistical inference must be continually tested and evaluated against their utility for applications, a point made strongly in his 2006 book (46) and again in 2015 (56).

Figure 5. With David Hinkley in the late 1970s. Their 1974 text (15) is a modern classic. (Courtesy of S. Stigler.)

His 2006 book, *Principles of statistical inference* (46), includes an extensive treatment of Bayesian theory and methods, which was timely, as Bayesian methods were being increasingly used in applications. The book includes an appendix titled ‘A personal view’, where he more explicitly insists on the necessity of an assessment of inferential methods based on their performance in a frequentist sense. In the index we find the entry ‘dispassionate assessment, frail attempt at, 1–296’; the appendix starts on page 297. In a note sent to Professor Anthony Edwards (FRS 2015) with a copy of his book, David wrote:

> I’m fairly cautious about the impact of the book in that it really is very cryptic indeed on key issues but we will see. In particular quite apart from the Bayesian stuff I have essentially discarded (not rejected) the Neyman-Pearson machinery in favour of Fisher’s original approach and I am sure this is the right route.

David’s understated sense of humour comes through in what was certainly a deliberate choice of parenthetical remark.

Applications and public service

David described himself as a ‘scientist, who happens largely to specialise in the use of statistics’ (Statistics Views 2014). Much of his theoretical and methodological work arose out of his interest in applications.

Lessons drawn from his applied experience were summarized in the opening article of the inaugural issue of *Annals of applied statistics* (47), and at greater length in two books.

Applied statistics (20), published in 1981 jointly with Joyce Snell, consists of a concise and highly illuminating discussion of general principles followed by a series of carefully chosen case studies. The later (2011) *Principles of applied statistics* (52) with Christl Donnelly is a distillation of a lifetime of experience. It ranges over many topics rarely included in similar texts, such as how units are chosen for observation or experiment, what it means to measure a quantity, the importance of dimensional analysis, and much more. All his general writing on applied statistics stresses both the 'desirability of an intimate union between subject-matter and statistical aspects of an investigation' (47) and that arriving at secure conclusions depends more on scientific good sense than on technical mastery of complex methods. One of his favourite discussion questions about a recently attended talk, usually uttered privately to a student or colleague, was 'Was it sensible?' (Keogh 2023).

Because David had very broad scientific interests, was generous with his time and could identify the statistical essence of a problem very quickly, he was very often asked to give advice on statistical analysis for particular applications. Many, perhaps most, of these efforts are not recorded in his list of publications. In addition to being widely consulted by colleagues and acquaintances, David was sought after for many expert panels and working groups in different contexts—when once asked what it took to be knighted for services to statistics, he replied 'serve on an infinite number of committees in zero time'.

In 1988 David chaired the Department of Health Working Group on HIV infection and AIDS, which produced a report on short-term predictions for England and Wales—the 'Cox report' (23)—and a special edition of *Philosophical Transactions of the Royal Society* on the supporting methodology (27–29). The incubation period, the interval between infection with HIV and diagnosis of AIDS, was essential for predicting the course of the epidemic, and David wrote several papers on its estimation, emphasizing that at the beginning of an epidemic data will be extremely incomplete. Similar methods were used for modelling the COVID19 pandemic in early 2020.

In 1997 the Department of the Environment, Food and Rural Affairs (DEFRA) established an independent scientific group (ISG) on cattle tuberculosis (TB), to which David was appointed, along with six other experts. Their final report (Independent Scientific Group on Cattle TB 2007) was delivered to DEFRA 10 years later; David was heard to say during that time 'badgers had taken over his life'. A key question was the role of badgers in spreading tuberculosis to cows and, while conflicting opinions were very strongly held, there was little scientific evidence available. The ISG designed a rigorous randomized controlled experiment over heterogeneous spatial areas to monitor the effect of badger culling on the spread of bovine TB; this led to the unanticipated result that incidence of TB adjacent to the culled area actually increased after culling and cast doubt on the effectiveness of culling policies (44, 49). The conclusions were made more secure by the careful design and analysis of the randomized controlled experiment. Christl Donnelly (FRS 2016) served on the ISG with David, and wrote (Donnelly 2023):

> Throughout, David was a source of wisdom and calm. Emotions sometimes ran high, but David's message to me was the same: 'Don't worry about the politics, the science will endure.'

A collaboration with the Liou laboratory at the University of Utah on cystic fibrosis (60) highlighted for David a concern that conventionally estimated standard errors were much too small when based on so-called big data, which he addressed from a theoretical point of view in (55) (figure 6).

Figure 6. David signed himself up as a speaker at his ninetieth birthday conference. His original and well-received talk on 'Big data' was published as (55). (Reproduced with the kind permission of the warden and fellows of Nuffield College, Oxford.)

John Goldthorpe has noted that 'rather little sociological research went on in [Nuffield] College to which David did not, in some way or other, make a significant contribution'. This included research on measuring the quality of life, on social mobility, on large-scale surveys, and more.

Further reading

In 1993 I interviewed David for a review journal in statistics (Reid 1994). He was a giant in the field, and had been for many years. Our conversation ranged over many of the topics described here. As I wrote elsewhere, neither of us could have predicted that he had nearly 30 productive years ahead of him (Reid 2023). Several of his publications in these later years were general overviews of statistical thinking, sometimes for specific application areas. To my mind, even more striking are a number of David's later papers that address a variety of current topics in statistical science. Examples include: the interpretation of very large numbers of two-sample comparisons, such as arise in gene expression data (43); a comment on models relevant for quantile regression (48); an alternative method for simulating from models with complicated likelihood functions (53); and an approach to model selection that identifies sets of models consistent with the observations (58, 59). These papers are a blend of classical ideas and modern problems, typically very short, with few references, but in a voice as fresh as that in his first research papers.

David published 23 books, most of them in a monograph series published by Methuen, and later Chapman & Hall; he also served as a long-time editorial advisor for the 'little green' series, so-called for the books' brevity and the colour of the covers. He said that it took him on average five years to write a book, but he must have meant from conception to publication, as he seemed to write very quickly. Often the monographs were prepared once he felt that he had captured a relatively broad view of a subject, and that he had something new to say about it. There is very often new material included, and not always flagged. I recall asking which paper underlay the discussion in §6.9 of his 2006 book (46), and he replied 'Well I just worked it out; it seemed to be needed'.

David was presented with a book of invited papers on a range of topics in statistics and probability to mark his sixty-sixth birthday (Hinkley *et al.* 1989). Another volume of invited papers followed the conference in honour of his eightieth birthday (Davison *et al.* 2005). The two-volume selected works (Hand & Herzberg 2006a, 2006b) reprints 86 papers published before 1993. Especially valuable are David's introductory comments to each paper, giving the historical context and connections to current research in statistical science.

A complete list of his publications is available online at https://doi.org/10.1098/rsbm.2023.0052.

Coda

When David started his career in statistics the field was fractious and scattered. Sir Ronald Fisher was famously quarrelsome, and estranged from many former colleagues. Jerzy Neyman (ForMemFRS 1979) held powerful sway at the University of California at Berkeley. There were bitter arguments at the RSS about the foundations of inference. Against this, David spoke and wrote succinctly, insightfully and calmly. He had very strong ideas about science and about statistics, yet his criticisms were unfailingly constructive. His originality and breadth shone from every page of his writing, and by the time he arrived in America in 1959, just 10 years from his PhD, he was widely recognized as an exceptionally original intellect.

In the brief abstract of his foundational paper (5) we read 'it consists of some general comments, few of them new, about statistical inference [...] Parts of the paper are controversial; these are not put forward in any dogmatic spirit'. This understated and self-effacing attitude was absolutely typical of David, and by virtue of his stature in the profession he raised the tone of our debates, and helped to make the community happier and more welcoming.

As news of David's death circulated around the world, tributes flooded social media and email in the community of statistical scientists. While many spoke of his intellectual achievements and his inspirational lectures, the majority commented on his generosity and humility. Memorial sessions were held at the Joint Statistical Meetings in Washington DC in August 2022, at the Royal Statistical Society in Aberdeen in September 2022 and at the London School of Hygiene and Tropical Medicine in November 2022. Sylvia Richardson, president of the RSS, introduced the Aberdeen session with (Richardson 2022):

> It is difficult to overstate just how important and influential Sir David Cox was to the community of statistical scientists, and how much he was respected, and cherished, by those who had the good fortune to work with, and learn from, him. The RSS described Sir David as 'one of the most important statisticians of the past century', and you would struggle to find anyone who would argue with that, except maybe Sir David himself. Because despite all his achievements in life, Sir David was always a thoroughly humble man.

Speakers and audience members at these occasions recalled times when David had replied to an email, a manuscript, a letter or a visit with extraordinary kindness and generosity, both of his time and his keen intellect, and described how their work and lives were the better for it.

A special section of *Significance* (Firth *et al.* 2022) prepared in the weeks immediately following his death includes personal tributes by several of his collaborators. A collection of reminiscences forms a special issue of the *Harvard Data Science Review* (Richardson & Wermuth 2023).

Awards and recognition

1961, 1973	Guy Medal in Silver, Gold, Royal Statistical Society
1973	Fellow of the Royal Society
1974	Foreign Honorary Member, American Academy of Arts and Sciences
1983	Foreign Member, Royal Danish Academy of Science and Letters
1985	Knighted
1988	Foreign Associate, National Academy of Sciences, USA
1990	Honorary Fellow, St John's College, Cambridge
1990	Kettering Prize and Gold Medal for Cancer Research
1992	Max Planck Forschungpreise
1996	Foreign Associate, Indian Academy of Sciences
1997	Honorary Fellow of the British Academy
2003	Honorary Fellow, American Academy of Political and Social Science
2005	George Box Medal
2010	Copley Medal
2013	Honorary Fellow, Royal Society of Edinburgh
2016	Inaugural International Prize in Statistics
2016	BBVA Foundation Frontiers of Knowledge Award

Acknowledgements

I am grateful to David's children for providing many family references and a detailed list of published tributes. David had an extensive network of colleagues who find themselves united by the common bond of friendship with this singular individual. While it was not possible to consult them all, I thank Heather Battey, Anthony Davison, Bradley Efron, David Firth, Valerie Isham, Peter McCullagh and Stephen Stigler for helpful comments.

The frontispiece portrait photograph is used with permission of the Jane Brown Estate.

Author profile

Nancy Reid FRS

Nancy Reid OC, FRS, FRSE, FRSC is a university professor in the Department of Statistical Sciences, University of Toronto. Her research field is statistical theory and foundations; her work in likelihood inference, especially the 1987 paper (21), has been widely cited. She held a NATO postdoctoral fellowship at Imperial College in 1979–1980, the happiest consequence of which was her collaboration and friendship with David Cox.

References to other authors

Battey, H. 2023 D. R. Cox. *Harvard Data Sci. Rev.* **5**(2). (https://hdsr.mitpress.mit.edu/pub/yiwo2meq) (doi:10.1162/99608f92.aedb3266)

Battey, H. & Reid, N. 2022 Obituary. *Bull. Inst. Math. Stat*. **51**, 14–16.

Birnbaum, A. 1962 On the foundations of statistical inference. *J. Am. Stat. Assoc.* **57**, 269–306. (doi:10.1080/01621459.1962.10480660)

Cambridge Statistical Laboratory. 2019 *Gallery of photographs 1953–2019*. Cambridge, UK: University of Cambridge. See https://www.statslab.cam.ac.uk/gallery-photographs-1953-2021.

Daniels, H. E. 1954 Saddlepoint approximations in statistics. *Ann. Math. Stat.* **25**, 631–650. (doi:10.1214/aoms/1177728652)

Daniels, H. E. 1958 Discussion of 'The regression analysis of binary sequences' by D. R. Cox. *J. R. Stat. Soc. B* **20**, 236–238.

Davison, A. C. 2023 Some reminiscences of David Cox. *Harvard Data Sci. Rev.* **5**(2). (doi:10.1162/99608f92.85038493)

Davison, A. C., Dodge, Y. & Wermuth, N. (eds). 2005 *Celebrating statistics: papers in honour of Sir David Cox on his 80th birthday*. Oxford, UK: Oxford University Press.

Davison, A. C., Isham, V. S. & Reid, N. 2022 Sir David Cox. *J. R. Stat. Soc. A* **185**, 2295–2306. (doi:10.1111/rssa.12964)

Donnelly, C. A. 2023 Sir David Cox. *Harvard Data Sci. Rev.* **5**(2). (doi:10.1162/99608f92.25212c73)

Efron, B. 2023 Remembering David Cox. *Harvard Data Sci. Rev.* **5**(2). (doi:10.1162/99608f92.87cef65e)

Firth, D., Reid, N., Mayo, D. G., Battey, H., Tibshirani, R., Borja, M. C., Stander, J. & Sebastiani, G. 2022 Profiles. *Significance* **19**, 30–41. (doi:10.1111/1740-9713.01632)

Hand, D. J. & Herzberg, A. M. (eds). 2006a *Selected statistical papers of Sir David Cox: volume 1, design of investigations, statistical methods and applications*. Cambridge, UK: Cambridge University Press.

Hand, D. J. & Herzberg, A. M. (eds). 2006b *Selected statistical papers of Sir David Cox: volume 2, foundations of statistical inference, theoretical statistics, time series and stochastic processes*. Cambridge, UK: Cambridge University Press.

Hinkley, D. V., Reid, N. & Snell, E. J. (eds). 1989 *Statistical theory and modelling: essays in honour of Sir David Cox*. London, UK: Chapman & Hall.

Independent Scientific Group on Cattle TB 2007 *Bovine TB: the scientific evidence*. See http://www.bovinetb.info/docs/final_report.pdf.

Isham, V. S. 2023 Memories of Sir David Cox. *Harvard Data Sci. Rev.* **5**(2). (doi:10.1162/99608f92.8d7d48fd)

Kalbfleisch, J. D. & Schnaubel, D. E. 2023 Fifty years of the Cox model. *Ann. Rev. Stat. Applic.* **10**, 1–23. (doi:10.1146/annurev-statistics-033021-014043)

Kartsonaki, C. 2023 David R. Cox. *Harvard Data Sci. Rev.* **5**(2). (doi:10.1162/99608f92.fe700a61)

Keogh, R. H. 2023 D. R. Cox. *Harvard Data Sci. Rev.* **5**(2). (doi:10.1162/99608f92.8e56c54f)

Levin, S. A. & Rinaldo, A. 2022 Ignacio Rodríguez-Iturbe (1942–2022). *Proc. Natl Acad. Sci.* **119**, e2217606119. (doi:10.1073/pnas.2217606119)

Reid, N. 1994 A conversation with Sir David Cox. *Stat. Sci.* **9**, 439–455. (doi:10.1214/ss/1177010394)

Reid, N. 2023 On 'A conversation with Sir David Cox'. *Harvard Data Sci. Rev.* **5**(2). (doi:10.1162/99608f92.81172cdd)

Richardson, S. 2022 *A celebration of the life and work of Sir David Cox*. See https://www.youtube.com/watch?v=ANHLgSE2dTw.

Richardson, S. & Wermuth, N. 2023 Remembering David Cox. *Harvard Data Sci. Rev*. **5**(2). (doi:10.1162/99608f92.35281d57)

Savage, L. 1954 *Foundations of statistics*. New York, NY: Dover.

Statistics Views. 2014 Interview with Sir David Cox. *Stats Data Sci. Views*. See https://www.statisticsviews.com/article/i-would-like-to-think-of-myself-as-a-scientist-who-happens-largely-to-specialise-in-the-use-of-statistics-an-interview-with-sir-david-cox/.

Valsecchi, M. G. 2014 Interview with Sir David Cox. International Biometric Society—Italian Region. See https://www.youtube.com/watch?v=lq8DrgkZlYE.

Van Noorden, R., Maher, B. & Nuzzo, R. 2014 The top 100 papers. *Nature* **514**, 550–553. (doi:10.1038/514550a)

Bibliography

The following publications are those referred to directly in the text. A full bibliography is available as electronic supplementary material at https://doi.org/10.6084/m9.figshare.c.7407194.

(1) 1948 (With A. Brearley) *An outline of statistical methods for use in the textile industry*. Leeds, UK: Wool Industries Research Association. (Six editions through to 1960.)

(2) 1949 The use of the range in sequential analysis. *J. R. Stat. Soc. B* **11**, 101–114. (doi:10.1111/j.2517-6161.1949.tb00027.x)

(3) 1955 Some statistical methods connected with series of events (with discussion). *J. R. Stat. Soc. B* **17**, 129–164. (doi:10.1111/j.2517-6161.1955.tb00188.x)

(4) 1958 *Planning of experiments*. New York, NY: Wiley.

(5) Some problems connected with statistical inference. *Ann. Math. Stat.* **29**, 357–372. (doi:10.1214/aoms/1177706618)

(6) The regression analysis of binary sequences (with discussion). *J. R. Stat. Soc. B* **20**, 215–242. (doi:10.1111/j.2517-6161.1958.tb00292.x)

(7) 1961 (With W. L. Smith) *Queues*. London, UK: Methuen.

(8) 1962 (With K. H. V. Booth) Some systematic supersaturated designs. *Technometrics* **4**, 489–495. (doi:10.1080/00401706.1962.10490035)

(9) *Renewal theory*. London, UK: Methuen.

(10) 1965 (With H. D. Miller) *The theory of stochastic processes*. London, UK: Methuen.

(11) 1966 (With P. A. W. Lewis) *The statistical analysis of series of events*. London, UK: Methuen.

(12) 1968 Notes on some aspects of regression analysis (with discussion). *J. R. Stat. Soc. A* **131**, 265–279. (doi:10.2307/2343523)

(13) 1970 *The analysis of binary data*. London, UK: Methuen.

(14) 1972 Regression models and life tables (with discussion). *J. R. Stat. Soc. B* **34**, 187–220. (doi:10.1111/j.2517-6161.1972.tb00899.x)

(15) 1974 (With D. V. Hinkley) *Theoretical statistics*. London, UK: Chapman & Hall.

(16) 1975 Partial likelihood. *Biometrika* **62**, 269–276. (doi:10.1093/biomet/62.2.269)

(17) 1979 (With O. E. Barndorff-Nielsen) Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. R. Stat. Soc. B* **41**, 279–312. (doi:10.1111/j.2517-6161.1979.tb01085.x)

(18) 1980 Local ancillarity. *Biometrika* **67**, 279–286. (doi:10.1093/biomet/67.2.279)

(19) 1981 Statistical analysis of time series. *Scand. J. Stat.* **8**, 93–115.

(20) (With E. J. Snell) *Applied statistics: principles and examples*. London, UK: Chapman & Hall.

(21) 1987 (With N. Reid) Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Stat. Soc. B* **49**, 1–39. (doi:10.1111/j.2517-6161.1987.tb01422.x)

(22) (With I. Rodríguez-Iturbe & V. Isham) Some models for rainfall based on stochastic point processes. *Proc. R. Soc. Lond. A* **410**, 269–288. (doi:10.1098/rspa.1987.0039)

(23) 1988 (With members of Working Party) *Short-term prediction of HIV infection and AIDS in England and Wales*. London, UK: HMSO.

(24) (With V. Isham) A simple spatial-temporal model of rainfall. *Proc. R. Soc. Lond. A* **415**, 317–328. (doi:10.1098/rspa.1988.0016)

(25) (With I. Rodríguez-Iturbe & V. Isham) A point process model for rainfall. *Proc. R. Soc. Lond. A* **417**, 283–298. (doi:10.1098/rspa.1988.0061)

(26) 1989 (With O. E. Barndorff-Nielsen) *Asymptotic techniques for use in statistics*. London, UK: Chapman & Hall.

(27) (With A. C. Davison) Prediction for small subgroups. *Phil. Trans. R. Soc. Lond. B* **325**, 185–187. (doi:10.1098/rstb.1989.0084)

(28) (With R. M. Anderson & H. C. Hillier) Epidemiological and statistical aspects of the AIDS epidemic. *Phil. Trans. R. Soc. Lond. B* **325**, 39–44. (doi:10.1098/rstb.1989.0073)

(29) (With G. F. Medley) A process of events with notification delay and the forecasting of AIDS. *Phil. Trans. R. Soc. Lond. B* **325**, 135–145. (doi:10.1098/rstb.1989.0078)

(30) 1989 (With E. J. Snell) *The analysis of binary data*, 2nd edn. London, UK: Chapman & Hall.

(31) 1993 (With N. Wermuth) Linear dependencies represented by chain graphs (with discussion). *Stat. Sci.* **8**, 204–283.

(32) 1994 (With O. E. Barndorff-Nielsen) *Inference and asymptotics*. London, UK: Chapman & Hall.

(33) (With V. Isham) Stochastic models of precipitation. In *Statistics for the environment. 2: water related issues* (eds V. Barnett & K. F. Turkman), pp. 3–19. Chichester, UK: Wiley.

(34) 1996 (With N. Wermuth) *Multivariate dependencies*. London, UK: Chapman & Hall.

(35) 1998 (With V. Isham) Stochastic spatial-temporal models for rain. In *Stochastic models in hydrology* (eds O. E. Barndorff-Nielsen, S. K. Gupta & E. Waymire), pp. 1–24. Singapore, Singapore: World Scientific.

(36) 1999 (With I. Rodríguez-Iturbe, A. Porporato, L. Ridolfi & V. Isham) Probabilistic modelling of water balance at a point. *Proc. R. Soc. Lond. A* **455**, 3789–3805. (doi:10.1098/rspa.1999.0477)

(37) 2000 (With N. Reid) *The theory of the design of experiments*. London, UK: Chapman & Hall/CRC Press.

(38) 2001 (With N. Wermuth) Causal inference and statistical fallacies. In *International encyclopedia of social and behavioural sciences* (eds N. J. Smelser & P. B Baltes), pp. 1554–1561. Oxford, UK: Elsevier.

(39) (With N. Wermuth) Some statistical aspects of causality. *Eur. Sociol. Rev.* **17**, 65–74. (doi:10.1093/esr/17.1.65)

(40) 2002 (With P. J. Solomon) *Components of variance*. London, UK: Chapman & Hall/CRC Press.

(41) 2003 Henry Ellis Daniels. *Biogr. Mems Fell. R. Soc.* **49**, 133–146. (doi:10.1098/rsbm.2003.0008)

(42) 2004 (With N. Wermuth) Causality. *Int. Stat. Rev.* **72**, 285–305. (doi:10.1111/j.1751-5823.2004.tb00237.x)

(43) (With M. Y. Wong) A simple procedure for the selection of significant effects. *J. R. Stat. Soc. B* **66**, 395–400. (doi:10.1111/j.1369-7412.2004.05695.x)

(44) 2005 (With others) Positive and negative effects of widespread badger culling on tuberculosis in cattle. *Nature* **439**, 843–846. (doi:10.1038/nature04454)

(45) 2006 (With V. Isham, I. Rodríguez-Iturbe, A. Porporato & S. Manfreda) Space-time modeling of soil moisture. *Water Resources Res.* **42**, 4497–4508. (doi:10.1029/2005WR004497)

(46) *Principles of theoretical inference*. Cambridge, UK: Cambridge University Press.

(47) 2007 Applied statistics. *Ann. Appl. Stat.* **1**, 1–18. (doi:10.1214/07-AOAS113)

(48) On a generalization of a result of W. G. Cochran. *Biometrika* **94**, 755–764. (doi:10.1093/biomet/asm046)

(49) (With others) Impacts of widespread badger culling on cattle tuberculosis. *Infect. Dis.* **11**, 300–308. (doi:10.1016/j.ijid.2007.04.001)

(50) 2010 (With D. Mayo) Frequentist statistics as a theory of inductive inference. In *Error and inference* (eds D. G. Mayo & A. Spanos), pp. 247–275. Cambridge, UK: Cambridge University Press.

(51) (With D. Mayo) Objectivity and conditionality in frequentist inference. In *Error and inference* (eds D. G. Mayo & A. Spanos), pp. 276–304. Cambridge, UK: Cambridge University Press.

(52) 2011 (With C. A. Donnelly) *Principles of applied statistics*. Cambridge, UK: Cambridge University Press.

(53) 2012 (With C. Kartsonaki) The fitting of complex parametric models. *Biometrika* **99**, 741–747. (doi:10.1093/biomet/ass030)

(54) 2014 (With R. Keogh) *Case-control studies*. Cambridge, UK: Cambridge University Press.

(55) 2015 Big data and precision. *Biometrika* **102**, 712–716. (doi:10.1093/biomet/asv033)

(56) (With N. Reid) On some principles of statistical inference. *Int. Stat. Rev.* **83**, 293–308. (doi:10.1111/insr.12067)

(57) 2016 Some pioneers of modern statistical theory. *Biometrika* **103**, 747–759. (doi:10.1093/biomet/asw052)

(58) 2017 (With H. S. Battey) Large numbers of explanatory variables, a semi-descriptive analysis. *Proc. Natl Acad. Sci.* **114**, 8592–8595. (doi:10.1073/pnas.1703764114)

(59) 2018 (With H. S. Battey) Large numbers of explanatory variables. *Proc. R. Soc. A* **474**, 20170631. (doi:10.1098/rspa.2017.0631)

(60) (With A. M. Granchelli, F. R. Adler, R. H. Keogh, C. Kartsonaki & T. G. Liou) Microbial interactions in the cystic fibrosis airway. *J. Clin. Microbiol*. **56**, 10.1128/jcm.00354-18. (doi:10.1128/JCM.00354-18)

(61) 2020 Statistical significance. *Ann. Rev. Stat. Appl*. **7**, 1–10. (doi:10.1146/annurev-statistics-031219-041051)