

Harvard Data Science Review • Issue 5.2, Spring 2023

D. R. Cox: Approach to Applied Statistics and the Importance of Being Sensible

Ruth H. Keogh¹

¹Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, England, United Kingdom

The MIT Press

Published on: Apr 27, 2023

DOI: <https://doi.org/10.1162/99608f92.8e56c54f>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

I first met David Cox in 2003, when he taught on the MSc in Applied Statistics at the University of Oxford. Over the summer of 2003 he was the supervisor for my master's degree research project, leading to me starting a DPhil with David as my supervisor, in which I focused on the design and analysis of case-control studies. This led to our book on the same topic (Keogh & Cox, 2014). David's influence on my statistical thinking and career is hard to overstate. Techniques that he developed—logistic regression (Cox, 1958) and the Cox model (Cox, 1972)—feature in my work on an almost-daily basis. I would like to focus here on some of the things I learned from David from his approach to applied statistics. From when I was a student, David was generous in exposing me to the applied work he was contributing to. I particularly remember learning about the research he was involved in on bovine tuberculosis and the role of badgers, with Professor Christl Donnelly and others. He introduced me to the field of cystic fibrosis, one of my main areas of research, through his collaborations with physician Dr. Theodore Liou (Liou, 2019).

In his review paper on applied statistics (Cox, 2007, p. 1) David wrote:

“What are the central principles of applied statistics? The great variety of the fields in which statistical considerations play some part makes it hazardous to attempt an answer. Any simple recommendation along the lines *in applications one should do so and so* is virtually bound to be wrong in some or, indeed, possibly many contexts. On the other hand, descent into a yawning abyss of vacuous generalities is all too possible.”

This reflects the approach that I saw him follow and recommend in practice. David was not keen on the idea that one must always do this or that or use such-and-such a method when faced with a particular question or challenge. He preferred a more pragmatic—or, as he might have said, ‘sensible’—approach. I think he was wary of relying too much on methods that make many assumptions or arrive too firmly at a single conclusion. Instead, he would encourage a more exploratory approach in which one gathers information about the extent to which certain features or assumptions have an important influence on what one might conclude from an analysis. In their book *Principles of Applied Statistics*, Cox and Donnelly (2011, p. 186) wrote:

“Judgement has to be exercised in how simple a model may be used, but it is often a good idea to start with a rather simple model and then to ask the question ‘What assumptions have been made that might seriously affect the conclusions?’ It may or may not be clear that no change in the model assumptions is likely seriously to change the conclusions.”

It seemed that David had a great ability to see applied problems clearly, and to break things down to the essential components, so that one could understand the issues by tackling them from a relatively simple starting point. In his comment on the paper of Breiman (2001) “Statistical Modelling: The Two Cultures,” David wrote: “If a simple standard model is adequate to answer the subject matter question, this is fine: there are severe hidden penalties for overelaboration” (p. 218). If we had been to a seminar to hear a talk, sometimes he would

say afterwards—“They made it so complicated!” Or if I was telling him about a talk I had attended, he would ask “Was it sensible?” I think that taking a ‘sensible’ approach to tackling a problem is something he placed a big emphasis on—recognizing the complications of a problem, but not making it more complicated than necessary. None of this is to say that David was conservative or excessively sensible, and he warned against being overcautious. In a talk to the Royal Statistical Society in 2014, for their 180th anniversary lecture series, David spoke on the topic of “Statistics – past, present and future” (Royal Statistical Society, 2020). I remember it well and [a recording is available](#). In that lecture he said: “Scientific research is not for the cautious. It’s an adventurous process. So how do we fulfill our role of caution and yet at the same time not go over the top in negativity. The answer must be something like the balance between analysis and interpretation” (RoyalStatSoc, 2014).

Although David seemed to have a great interest in applied problems, it was also with a view on how problems arising in a specific example were representative of more general challenges. In his 1994 interview with Nancy Reid, he said: “The things that are going to be most widely useful are those where you stand back from one very particular application and say here’s a whole family of problems that arise in applications in several fields and try to address that” (Reid, 1994, p. 451). Some of David’s recent work developed approaches that enable exploratory or semi-descriptive analyses in response to common challenges, including for problems involving large numbers of explanatory variables relative to the number of individuals (Cox & Battey, 2017) and for data involving missing values (Battey & Cox, 2023).

David was enormously kind and generous to me, as supervisor, colleague, and friend. We shared many lunches together at Nuffield College in Oxford, often on a Friday when fish and chips were being served (he would always request the “smallest piece of fish, please,” though they were all pretty huge). We had a nice time traveling to talks or meetings together, at which he was usually more interested in hearing about the work of more junior researchers rather than the big keynote speaker, reflecting his encouragement of young people. I am very fortunate to have had the chance to learn from David, about statistics and beyond. I think that the Royal Statistical Society (2020) put it perfectly when they said: “As well as his immense contributions to statistical research, Sir David will be remembered as a boundlessly generous and supportive friend to generations of statisticians. His kindness and humility were as remarkable as his genius.”

Disclosure Statement

Ruth H. Keogh has no financial or non-financial disclosures to share for this article.

References

Battey H. S., & Cox, D. R. (2023). Missing observations in regression: A conditional approach. *Royal Society Open Science*, 10(2), Article 220267. <https://doi.org/10.1098/rsos.220267>

Cox, D. R. (2007). Applied statistics: A review. *Annals of Applied Statistics*, 1(1), 1–16.

<https://doi.org/10.1214/07-AOAS113>

Cox, D. R. (2001). Comment on “Breiman, L. (2001). Statistical modelling: The two cultures. *Statistical Science*, 16, 199–231.” *Statistical Science*, 16(3), 216–218.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society (Series B)*, 34(2), 187–202. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society (Series B)*, 20(2), 215–232. <https://doi.org/10.1111/j.2517-6161.1958.tb00292.x>

Cox, D. R., & Donnelly, C. A. (2011). *Principles of applied statistics*. Cambridge University Press.

<https://doi.org/10.1017/CBO9781139005036>

Cox, D. R., & Battey, H. S. (2017). Large numbers of explanatory variables, a semi-descriptive analysis. *Proceedings of the National Academy of Sciences*, 114(32), 8592–8595.

<https://doi.org/10.1073/pnas.1703764114>

Keogh, R. H., & Cox, D. R. (2014). *Case-control studies*. Cambridge University Press.

<https://doi.org/10.1017/CBO9781139094757>

Liou, T. G., Adler, F. R., Argel, N., Asfour, F., Brown, P. S., Chatfield, B. A., Daines, C. L., Durham, D., Francis, J. A., Glover, B., Heynekamp, T., Hoidal, J. R., Jensen, J. L., Keogh, R., Kopecky, C. M., Lechtzin, N., Li, Y., Lysinger, J., Molina, O., . . . Cox, D. R. (2019); Prospective multicenter randomized patient recruitment and sample collection to enable future measurements of sputum biomarkers of inflammation in an observational study of cystic fibrosis. *BMC Medical Research Methodology*, 19(1), Article 88.

<https://doi.org/10.1186/s12874-019-0705-0>

Reid, N. (1994). A conversation with Sir David Cox. *Statistical Science*, 9(3), 439–455.

<https://doi.org/10.1214/ss/1177010394>

Royal Statistical Society. (2020, January 22). *Sir David Cox, 1924-2022*. <https://rss.org.uk/news-publication/news-publications/2022/general-news/sir-david-cox-1924-2022/>

RoyalStatSoc. (2014, Apr 10). *Sir David Cox: Statistics - past, present and future* [Video]. YouTube.

<https://www.youtube.com/watch?v=xRik3vOKLcU>

©2023 Ruth H. Keogh. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.

