CC/NUMBER 42
OCTOBER 20, 1986

This Week's Citation Classic®

Cox D R. Regression models and life-tables.
J. Roy. Statist. Soc. Ser. B Metho. 34:187-220, 1972.
[Imperial College, London, England]

The paper concerned the analysis of a common type of failure (or survival) data in which the dependence on explanatory variables is studied. Part of the investigation concerned the setting up of suitable models and part with the methods for detailed statistical analysis. [The *Science Citation Index*® (*SCI*®) and the *Social Sciences Citation Index*® (*SSCI*®) indicate that this paper has been cited in over 2,095 publications.]

David Cox
Department of Mathematics
Imperial College
London SW7 2AZ
England

August 7, 1986

One of the challenges in discussing work done in the remote past is to avoid unconscious self-justificatory hindsight. So far as I now recall, four things led to the paper.

The first two were an interest in industrial life testing and associated statistical problems and a corresponding interest in probability models for series of point events.

The third and crucial stimulus was that much survival data was being collected, especially in clinical trials, in which there were many potential explanatory variables and in which analysis and interpretation had to be made when, for some individuals, only a lower limit to survival time is known. There was a need for flexible methods of analysis. I will come to the fourth consideration in a moment.

The first step was to formulate a number of potentially useful general families of statistical models. One was suggested fairly naturally from ideas in stochastic processes. Namely, one should consider the hazard or probability of immediate failure, given survival up to the time in question. By postulating provisionally a product form for the hazard in which dependence on the explanatory variables could be introduced as one of the factors, a suitable family of models could be written down and given a clear physical interpretation.

The next step—to produce a simple method of statistical analysis of empirical data—is where the fourth consideration, an interest in the general theory of statistical inference, ought, perhaps, to have given a speedy answer. In fact, the standard approach of writing down the likelihood function gave only an expression of virtually useless complexity. It took some years of thought (on and off) to come to the sudden realization that most of the likelihood was irrelevant to the main purpose and that if only relevant factors were retained, quite simple procedures of analysis were achieved. The new likelihood function corresponded to a clearly defined, although rather complex, series of conditioning operations, and I called it *a* conditional likelihood.

This and other material was presented to the Royal Statistical Society in 1972, the printed version being followed by details of the ensuing discussion. There was some objection to the term "conditional likelihood" because it was not *the* conditional likelihood as defined by others. I later introduced the term "partial likelihood" to avoid misunderstanding and also sketched a "proof" that the required properties of the estimates hold.[1]

Although I felt that a proof would have produced a theorem having confusing regularity conditions of no scientific interest, subsequent work by others on the more mathematical aspects of these and related problems has produced some strikingly elegant results.[2] Remarkably quickly, some of the producers of statistical packages included the method in their libraries, and this was the point at which the method began to be used on a fairly large scale.

By far the most satisfying development in this type of work has been the realization that the broad idea sketched in the 1972 paper is applicable in many fields of study ranging from econometrics to animal breeding. It is this potential that gives statistical theory and method one of its special attractions as a field in which to work.

[For an up-to-date review, see reference 3.]

1. Cox D R. Partial likelihood. *Biometrika* 62:269-76, 1975. (Cited 175 times.)
2. Gill R D. Understanding Cox's regression model: a martingale approach *J. Amer. Statist. Assn.* 79:441-7, 1984.
3. Cox D R & Oakes D. *Analysis of survival data.* London: Chapman & Hall, 1984. 201 p.

