Lifetime Data Anal
DOI 10.1007/s10985-007-9052-2

Norman Breslow, an architect of modern biostatistics

Nick E. Day · Mitchell H. Gail

Received: 17 July 2007 / Accepted: 30 July 2007

Since the early 1970's, Norman Breslow has made enormous contributions to statistical theory and its applications in epidemiology, and he was a founding member and principal statistician of the Wilms Tumor Study Group, which is credited with great improvements in the treatment and understanding of this disease. Applications of the biostatistical methods that Norm developed have been concentrated on two of the most important areas of biomedical science where statistical thinking is fundamental, clinical trials and chronic disease epidemiology. In these two areas, much of the applied statistical work over the past three decades is based on the conceptual framework of which he was a principal architect.

Biostatistics is a pluripotential discipline with a multitude of possible career avenues. Many biostatisticians loosen their ties with the core discipline to become, for example, epidemiologists, or to manage clinical trial data centers or move into population genetics, to name just some of the options. This Norm never did. Although he became deeply involved in a range of biomedical applications, in particular his extremely productive commitment over the decades to the Wilms Tumor Study Group, described in the later article by Giulio D'Angio,

N. E. Day
Epidemiology, University of Cambridge, Cambridge, UK

M. H. Gail (✉)
Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
e-mail: gailm@epndce.nci.nih.gov

Norm remained focused throughout his career on the development of the core statistical discipline. In the deepest traditions of scholarship, Norm's efforts entailed not only personal research, but also a commitment, described later by Polly Feigl, to training the next generations of practicing biostatisticians, to nurturing those in the early stages of a research career, and to advancing the international status of the profession. In this latter respect, his election as President of the International Biometrics Society was very broadly welcomed, and his presidential address in 2002 gave voice to what many feel, that the major international prize-giving bodies, notably the Nobel committee, have done biomedicine a great disservice by not properly recognizing the huge contribution that biostatistics and allied disciplines have made to improving the world's health over the last six or seven decades.

The 1970's saw the foundations of survival analysis completely recast. Norm was a leader in this area, giving rigorous understanding to the stochastic theory of the Kaplan–Meier curve (Breslow and Crowley 1974) and to the generalized Kruskal–Wallis test (Breslow 1970) of the equality of survival curves. In a series of studies, he developed and applied the proportional hazards assumption in epidemiology and helped create a new conceptual framework and set of tools for epidemiologists. One paper derived the properties of indirect standardization from a regression perspective (Breslow and Day 1975); another gave a proper statistical interpretation of the epidemiologist's "expected" number of events as a random quantity (Breslow 1975). A seminal paper (Prentice and Breslow 1978) established the connection between hazard ratios in prospective cohort studies and the relative odds in time-matched ("nested") case-control studies. A closely related paper on conditional logistic analysis of matched case-control data revolutionized the approach to analysis of such data (Breslow et al. 1978), as did a paper on unconditional logistic regression analysis of the log odds ratio (Breslow 1976). Later papers on design showed how efficient the nested case-control design could be, compared to the full cohort analysis (Breslow et al. 1983).

From my (ND) own viewpoint as a statistical epidemiologist, the development of these approaches was greatly liberating. For those who were not involved with the generation or analysis of epidemiological data in the mid 1970's, it may be useful to recall the situation. The tools available were limited in scope and not perceived to have any underlying unity. There had been little conceptual advance since the classic Mantel-Haenszel paper of 1959 (Mantel and Haenszel 1959), with the extension to trend tests (Mantel 1963). Multivariate estimates of risk, the simultaneous control of several confounding variables, and examination of interaction effects could only be dealt with by rather unsatisfactory ad hoc approaches. For unmatched studies, unconditional logistic regression was slowly being introduced, but for matched studies nothing was available. For anything more complex than a univariate analysis with categorized data, the matching had to be ignored. The workshop that Norm and I held at the IARC in December 1977 brought together practicing epidemiologists and biostatisticians to discuss the scope of the monograph on the analysis of case-control studies we were preparing. Software was available at the workshop for the analysis of matched studies using conditional logistic regression. The response from the workshop participants was clear, that despite some off-stage noises from the foggy northwest Atlantic, the analytic tools becoming available, based on unconditional and conditional logistic regression, provided a complete answer to the analytic questions being asked of data at the time. Not only did they offer new and more flexible analytic approaches, but they also provided a framework to motivate and assess the simpler contingency table methods, such as those of Mantel and Haenszel. No living discipline stands still, but there was a feeling that a stage had been completed in the evolution of the subject. This is perhaps born out by the fact that more than 25 years after publication of the ensuing monograph (Breslow and Day 1980), this book was still the most quoted reference in the American Journal of Epidemiology. Seven years later, the complementary publication on cohort studies (Breslow and Day 1987) appeared, which in a logical world might have been Volume 1 rather than Volume 2.

This has been a very selective summary of some of Norm's statistical contributions, chosen from a bibliography containing more than 200 refereed articles and more than invited 70 chapters and other writings. Among other statistical research themes that should be mentioned are: development of a satisfactory asymptotic theory and methods for the analysis of sparse data, as in matched case-control studies analyzed by the Mantel-Haenszel method (Breslow 1981; Breslow and Liang 1982; Robins et al. 1986); allowance for overdispersion in Poisson data (Breslow 1984,1990); and random effects modeling in general linear models (Breslow and Clayton 1993; Lin and Breslow 1996). Another important research theme is the design and analysis of two-phase studies, such as two-phase case-control studies in which readily available covariates are obtained on all cases and controls, but hard to gather measurements are obtained selectively among cases and controls (Cain and Breslow 1988; Breslow and Holubkov 1997a, b; Chatterjee et al. 2003). This work is being used to increase the statistical efficiency of epidemiologic studies and to provide a framework for analysis when some of the case-control data are missing (Chen et al. 2006).

Norman Breslow has received many honors for his work, including the Snedecor and Fisher awards from the Committee of Presidents of Statistical Societies and an Honorary Doctorate from the University of Bordeaux. He is an Honorary Fellow of the Royal Statistical Society, a member of the Institute of Medicine of the National Academy of Sciences and a former President and Honorary Life Member of the International Biometric Society. Perhaps his greatest honor is the widespread adoption of the many conceptual approaches and analytic tools he has given us.

References

Breslow N (1970) A generalized Kruskal-Wallis test for comparing K samples subject to unequal amounts of censorship. Biometrika 57:579–594

Breslow NE (1975) Analysis of survival data under the proportional hazard model. Rev Int Stat 43:45–58

Breslow NE (1976) Regression analysis of the log odds ratio: a method for retrospective studies. Biometrics 32:409–416

Breslow N (1981) Odds ratio estimators when the data are sparse. Biometrika 68:73–84

Breslow NE (1984) Extra-Poisson variation in log-linear models. Appl Stat 33:38–44

Breslow NE (1990) Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. JASA 85:565–571

Breslow NE, Clayton DC (1993) Approximate inference in generalized linear mixed models. JASA 88:9–25

Breslow N, Crowley J (1974) A large sample study of the life table and product limit estimates under random censorship. Ann Stat 2:437–453

Breslow NE, Day NE (1975) Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. J Chron Dis 28:289–303

Breslow NE, Day NE (1980) Statistical methods in cancer research I: the analysis of case-control studies. International Agency for Research on Cancer, Lyon p 338

Breslow NE, Day NE (1987) Statistical methods in cancer research II: the design and analysis of cohort studies. International Agency for Research on Cancer, Lyon p 446

Breslow NE, Holubkov R (1997a) Weighted likelihood, pseudolikehood and maximum likelihood methods for logistic regression analysis of two-stage case-control data. Stat Med 16:103–116

Breslow NE, Holubkov R (1997b) Maximum likelihood estimation of logistic regression parameters under two-phase, outcome-dependent sampling. J R Stat Soc Ser B 59(2):447–461

Breslow NE, Liang KY (1982) The variance of the Mantel-Haenszel estimator. Biometrics 38:943–952

Breslow NE, Day NE, Halvorsen KT, Prentice RL, Sabai C (1978) Estimation of multiple relative risk functions in matched case-control studies. Am J Epidemiol 108:299–307

Breslow N, Lubin J, Marek P, Langholz B (1983) Multiplicative models and cohort analysis. JASA 78:1–12

Cain KC, Breslow NE (1988) Logistic regression analysis and efficient design for two-stage studies. Am J Epidemiol 128(6):1198–1206

Chatterjee N, Chen Y-H, Breslow NE (2003) A pseudoscore estimator for regression problems with two phase sampling. J Am Stat Assoc 98:158–168

Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, Benichou J, Gail MH (2006) Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst 98(17):1215–1226

Lin X, Breslow NE (1996) Bias correction in generalized linear mixed models with multiple components of dispersion. J Am Stat Assoc 91:1007–1016

Mantel N (1963) Chi-square tests with 1 degree of freedom-extensions of Mantel-Haenszel procedure. J Am Stat Assoc 58(303):690–700

Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 22(4):719–748

Prentice RL, Breslow NE (1978) Retrospective studies and failure time models. Biometrika 65:153–158

Robins J, Breslow N, Greenland S (1986) Estimators of the Mantel-Haenszel variance consistent in both sparse data and large strata limiting models. Biometrics 42:311–323