Preface

Jianwen Cai · Li Hsu

Received: 16 August 2013 / Accepted: 25 September 2013 / Published online: 16 October 2013
© Springer Science+Business Media New York 2013

1 Preface

Ross Prentice has made enormous contributions to the fields of biostatistics, epidemiology, and disease prevention, including survival analysis, case–cohort designs and the use of surrogate endpoints in clinical trials, nutritional epidemiology, genetic epidemiology, biomarkers and measurement error. He champions the interdisciplinary sciences and played central role in the conception, design and implementation of the Women’s Health Initiative, for which many of his methodological works in all three fields (biostatistics, epidemiology and disease prevention) converge. It is impossible to summarize all of these accomplishments in one issue, so this special issue of *Lifetime Data Analysis* in honor of Ross will focus on various survival analyses-related topics. [Oakes \(2013\)](#) provides an excellent expository review paper on many topics that Ross has made fundamental contributions to. Other areas where Ross has made substantial contributions, including the Women’s Health Initiative, measurement error, nutritional epidemiology, and genetic epidemiology, are covered in a special issue of *Statistics in Biosciences*.

Ross has authored many influential papers associated with clinical trials, including work on life history and failure time data. He defined a criterion for the use of surrogate endpoints in clinical trials ([Prentice 1989](#)). This criterion essentially requires the surrogate variable to “capture” the relationship (if any) between the treatment and the true endpoint. In the addendum, Jay Herson describes the humble beginning of this highly influential notion, now known as the “Prentice Criterion.” In another example, [Prentice et al. \(1981\)](#) developed partially conditional models for recurrent event

J. Cai (✉)

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
e-mail: cai@bios.unc.edu

L. Hsu

Biostatistics and Biomathematics Program, Fred Hutchinson Cancer Research Center,
Seattle, WA 98109, USA

data. This special issue contains several papers on clinical trials, exemplifying Ross' influence not only on the field, but also on the research and career development of some of the authors. [Diao et al. \(2013\)](#) extend modeling of recurrent events to include modeling of markers to reflect possible severity of the event, associated cost, or short term response to treatment. They describe a novel model for a marked point process, which incorporates dependence between continuous markers and the event process through the use of a copula function. [He et al. \(2013\)](#) present their work on evaluation of hospital readmission among kidney dialysis facilities in the United States, and they provide a novel algorithm for handling large numbers of parameters associated with facilities and hospitals. [Daniel and Tsiatis \(2013\)](#) consider the problem of estimating the distribution of time to composite endpoint when some endpoints are only partially observed. They propose incorporating the partial information through the augmented inverse probability weighted estimating equations to increase the efficiency of the estimator without relying on additional assumptions beyond those that would be made by standard approaches.

In his recent work Ross has also made major contributions to risk prediction models ([Wacholder et al. 2010](#); [Mealiffe et al. 2010](#)). In this issue of LIDA, [Cai et al. \(2013\)](#) propose a systematic approach to identify those individuals, using their conventional risk factors/markers, who would benefit from a new set of risk markers for more accurate classification. [Zheng et al. \(2013\)](#) study in depth a novel two-phase study design that involves sampling until quotas of eligible cases and controls are identified for validating biomarkers. This practical design could potentially save costs by not requiring assessment of all cohort members prior to phase II sampling. They develop semi-parametric methods to calculate risk distributions and a wide variety of prediction indices for time-to-event data under this design.

Large cohort studies or clinical trials that monitor individuals over time for disease occurrences are very powerful for studying multiple outcomes, estimating absolute incidence rates in exposed and non-exposed people. Because exposures are assessed before disease onset, there is no ambiguity about whether or not exposure precedes disease. However, these studies are costly to conduct. Ross proposed a cost-effective case-cohort study design. This involves a random sample of the entire cohort, called the subcohort, augmented with subjects having the disease of interest but not included in the subcohort sample. In contrast to the popular nested case-control study design, the case-cohort design allows multiple studies of different diseases to be conducted using the same subcohort. If the subcohort is of sufficient size, the statistical power for studying the association of risk factors with disease is not substantially reduced compared to the impractical alternative of collecting risk factors for the entire cohort. As Ross indicated, standard methods of statistical analysis for relative risk model require some modification for the case-cohort design. Ross formally formulated the design and proposed the analysis method based on a pseudo-likelihood function in 1986 ([Prentice 1986](#)) and later provided the theoretical foundation ([Self and Prentice 1988](#)). During the past 20 years, this study design has been widely used in large cohort studies such as the Women's Health Initiative (WHI Study Group 1998), in which Ross played a central role, and other large cohort studies such as the Atherosclerosis Risk in Communities (ARIC) study (The ARIC Investigators 1989), the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study ([Howard et al.](#)

2005), and the Busselton Health Study (Cullen 1972). In this special issue, Chen and Chen (2013) consider the case-cohort design for recurrent events with certain specific clustering features. They propose a properly modified Cox-type self-exciting intensity model. In the paper by Lin (2013), we see that the case-cohort design remains very useful in the current era of genetic studies.

Multivariate survival analysis is another area on which Ross has spent much time and effort. For the last two decades he has worked on the difficult topic of nonparametric estimation of the multivariate survival distribution. While van der Laan (1996) showed that asymptotically efficient nonparametric estimation of multivariate survival function is possible, further development is still needed for these estimators to be used in moderate sample sizes. Examples of his work in this area are Prentice and Cai (1992); Prentice (1999) and Fan et al. (2000). This special issue includes papers about several topics on multivariate survival times, including cross-ratio estimation for bivariate failure times with left truncation by Hu et al. (2013); a fixed-effects model for survival analysis in matched pairs studies by Gerster et al. (2013); and multi-state models for a wide variety of medical applications characterized by multiple events and longitudinal data (Farewell and Tom 2013).

This is a very selective summary of a few of Ross's many major statistical contributions in the area of survival analysis. Other themes in his research in this area include: development of classical methods for handling missing and mis-measured covariate data (e.g., Prentice 1982); application of counting process theory to survival models (Self and Prentice 1988); analysis of competing risks data (e.g., Prentice et al. 1978); development of linear rank tests with censored data (e.g., Prentice 1978); use of marginal models for correlated failure time data (e.g., Cai and Prentice 1995 and Prentice and Hsu 1997 that we have co-authored); development of failure time methods for statistical genetics and for gene and environment interaction (e.g., Zhong and Prentice 2010; Prentice et al. 2010). We thank the Editor-in-Chief for this opportunity to honor Ross for his numerous contributions, profound impact, and broad influence on statistical methodology and application, administrative leadership, and mentorship. Our personal experiences as Ross' students are invaluable. From Ross we learned to appreciate both statistical methodology and applications. Besides research, he showed us how to multi-task and take leadership responsibilities by his own actions. We owe him special thanks for his inspiration and guidance but we know that his influence extends far beyond his direct collaborators. Because of the enthusiastic responses, we received many interesting articles and hence, due to page limit, could not include all manuscripts in one issue. The papers on case-cohort design (Chen and Chen 2013; Lin 2013) and multivariate survival analysis (Hu et al. 2013; Gerster et al. 2013; Farewell and Tom 2013) will appear in the next issue.

2 Addendum: The humble beginnings of the prentice criterion

by Jay Herson, Ph.D.

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health,
Baltimore, MD, USA

Nearly everyone who has contemplated surrogate endpoints in clinical trials has encountered what has become known as the Prentice Criterion. However few people

may know how Ross came to work on this problem. Back in 1982, Steve George, then at St. Jude's Children's Hospital, invited me to be a discussant at a Society for Clinical Trials session on how to make clinical trials more efficient. Steve offered a list of possibilities and one of them was surrogate endpoints. I had an idea of what was meant by this term but did not include it in my discussion. Nevertheless, I was intrigued by this notion so when asked to organize a session for the 1987 joint ENAR-WNAR meeting in Dallas, TX I thought of surrogate endpoints. I invited Janet Wittes to speak on surrogate endpoints used in cardiovascular disease, Susan Ellenberg in oncology and Argye Hillis on ophthalmologic diseases. I was in several telephone discussions with Ross at this time regarding his calculation of sample size for the Women's Health Initiative lipids study so, on one of these occasions, I asked Ross if he would be discussant for this session and he accepted.

I hadn't heard from Ross at all in the time leading up to the meeting. Ross already had a reputation for being a brilliant theoretical and applied biostatistician and this caused considerable excitement and some angst among the presenters. Susan Ellenberg, in particular, was nervous about what Ross might say about her paper. Well, she was just a kid at the time.

When the day came I began the session by saying that I had been thinking of surrogate endpoints in clinical trials since 1982. The three presenters followed and then Ross took the podium. He began by saying "I have been thinking of surrogate endpoints since Jay called me to be discussant at this session." He then proceeded to present the Prentice Criterion for the first time and never did refer to any of the papers presented.

This criterion rapidly passed through the statistical, medical and regulatory literature. Whenever I encounter it I wonder where would clinical trial methodology stand if I had asked someone else to be discussant at that session.

References

- Cai T, Tian L, Lloyd-Jones D, Wei LJ (2013) Evaluating subject-level incremental values of new markers for risk classification rule. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9272-6](https://doi.org/10.1007/s10985-013-9272-6)
- Cai J, Prentice RL (1995) Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* 82:151–164
- Chen F, Chen K (2013) Case-cohort analysis of clusters of recurrent events. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9275-3](https://doi.org/10.1007/s10985-013-9275-3)
- Cullen KJ (1972) Mass health examinations in the Busselton population, 1966 to 1970. *Med J Aust* 2:714–718
- Daniel R, Tsiatis AA (2013) Efficient estimation of the distribution of time to composite endpoint when some endpoints are only partially observed. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9261-9](https://doi.org/10.1007/s10985-013-9261-9)
- Diao L, Cook RJ, Lee KA (2013) A copula model for marked point processes. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9259-3](https://doi.org/10.1007/s10985-013-9259-3)
- Fan J, Prentice RL, Hsu L (2000) A class of weighted dependence measures for bivariate failure time data. *J R Stat Soc Ser B* 62:181–190
- Farewell V, Tom BDM (2013) The versatility of multi-state models for the analysis of longitudinal data with unobservable features. *Lifetime Data Anal.* doi:[10.1007/s10985-012-9236-2](https://doi.org/10.1007/s10985-012-9236-2)
- Gerster M, Madsen M, Andersen P (2013) Matched survival data in a co-twin control design. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9256-6](https://doi.org/10.1007/s10985-013-9256-6)
- He, K., Kalbfleisch, J. D., Li, Y., & Li, Y. (2013). Evaluating hospital readmission rates in dialysis facilities; adjusting for hospital effects. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9264-6](https://doi.org/10.1007/s10985-013-9264-6).

- Howard VJ, Cushman M, Pulley L, Gomez C, Go R, Prineas RJ et al (2005) The REasons for Geographic And Racial Differences in Stroke (REGARDS) Study: objectives and design. *Neuroepidemiology* 25:135–143
- Hu T, Lin X, Nan B (2013) Cross-ratio estimation for bivariate failure times with left truncation. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9263-7](https://doi.org/10.1007/s10985-013-9263-7)
- Lin DY (2013) Survival analysis with incomplete genetic data. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9262-8](https://doi.org/10.1007/s10985-013-9262-8)
- Mealiffe ME, Stokowski RP, Rhee BK, Prentice RL, Pettinger M, Hinds DA (2010) Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *J Natl Cancer Inst* 102:1618–1627
- Oakes, D. An introduction to survival models: in honor of Ross Prentice. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9276-2](https://doi.org/10.1007/s10985-013-9276-2).
- Prentice RL (1978) Linear rank tests with right censored data. *Biometrika* 65(1):167–179
- Prentice RL (1986) A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 73:1–11
- Prentice RL (1989) Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 8(4):431–440
- Prentice RL (1982) Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 69:331–342
- Prentice RL (1999) On non-parametric maximum likelihood estimation of the bivariate survivor function. *Stat Med* 18(17–18):2517–2527
- Prentice RL, Cai J (1992) Covariance and survival function estimation using censored multivariate failure time data. *Biometrika* 79:495–512
- Prentice RL, Hsu L (1997) Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika* 84(2):349–364
- Prentice RL, Huang Y, Hinds DA, Peters U, Cox DR, Beilharz E et al (2010) Variation in the FGFR2 gene and the effect of a low-fat dietary pattern on invasive breast cancer. *Cancer Epidemiol Biomark Prev* 19:74–79
- Prentice RL, Kalbfleisch JD, Peterson AV Jr, Flournoy N, Farewell VT, Breslow NE (1978) The analysis of failure times in the presence of competing risks. *Biometrics* 34:541–554
- Prentice RL, Williams BJ, Peterson AV (1981) On the regression analysis of multivariate failure time data. *Biometrika* 68:373–379
- Self SG, Prentice RL (1988) Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann Stat* 16:64–81
- The ARIC Investigators (1989) The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. *Am J Epidemiol* 129:687–702
- Wacholder S, Hartge P, Prentice RL, Garcia-Closas M, Diver WR et al (2010) Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* 362:986–993
- Women's Health Initiative Study Group (1998) Design of the women's health initiative clinical trial and observational study. *Control Clin Trials* 19(1):61–109
- van der Laan MJ (1996) Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann Stat* 24(2):596–627
- Zheng Y, Cai T, Pepe MS (2013) Adopting nested case-control quota sampling designs for the evaluation of risk markers. *Lifetime Data Anal.* doi:[10.1007/s10985-013-9270-8](https://doi.org/10.1007/s10985-013-9270-8)
- Zhong H, Prentice RL (2010) Correcting 'winner's curse' in odds ratios from genomewide association findings for major complex human diseases. *Genet Epidemiol* 34(1):78–91