Editorial

Statistical Methods in Medical Research

Wenyaw Chan
University of Texas, Health Science Center at Houston, TX, USA

Since John Snow first conducted a modern epidemiological study in 1854 during a cholera epidemic in London, statistics has been associated with medical research. After Austin Bradford Hill published a series of articles on the use of statistical methodology in medical research in 1937, statistical considerations and computational tools have been paramount in conducting medical research [1]. For the past century, statistics has played an important role in the advancement of medical research and medical research has stimulated rapid development of statistical methods. For example, the development of modern survival analysis-an important branch of statistics has aimed to solve problems encountered in clinical trials and large-scale epidemiological studies. In this era of evidence-based medicine, the development of novel statistical methods will continue to be crucial in medical research. With the expansion of computer capacity and advancement of computational techniques, it is inevitable that modern statistical methods will likely incorporate, to a greater degree, complex computational procedures.

This issue focuses on statistical methods in medical research. Several novel methods aiming on solving different medical research questions are introduced. Some unique approaches of statistical analysis are also present.

Hanagal and Sharma contribute two papers. The first one deals with a bivariate survival model. They examine a parameter estimation issue when the samples are taken from a bivariate log-logistic distribution with shared gamma frailty. They propose to use a Bayesian approach along with the Markov Chain Monte Carlo computational technique for implementation. The computer simulation is conducted for performance evaluation. Two well-known datasets, one about acute leukemia and the other about kidney infection are applied as examples.

The second paper contributed by Hanagal and Sharma examines the shared inverse Gaussian frailty model with the bivariate exponential baseline hazard. They first derive the likelihood of the joint survival function. In their Bayesian approach, the parameters of the baseline hazard are assumed to follow a gamma distribution while the coefficients of the regression relationship are assumed to follow an independent normal distribution. The dependence of two components of the survival function is tested. Three information criteria are used for model comparisons. The proposed method is applied to analyze diabetic retinopathy data.

The paper by Chang, Iyer, Bullitt and Wang provides a method to find determinants of the brain arterial system. They represent the brain arterial system as a binary tree and apply the mixed logistic regression model to find significant covariates. The authors also demonstrate model selection methods for both fixed and random effects. A case study is presented using the method. This paper provides a rigorous approach for analyzing the binary branching structure data. It is potentially applicable to other tree structure data.

Chakraborty proposes two probabilistic models to estimate male-to-female HIV-1 transmission rate in one sexual contact. One model is applicable when the transmitter cell counts are known and the other model is applicable when the receptor cell counts are known. By first uniformizing each transmitter (or receptor) cell count and assuming as a beta distribution, this paper algebraically derives the transition probability by imposing some boundary conditions based on scientific phenomena related to HIV infection.

The paper by Yeh, Jiang, Garrard, Lei and Gajewski proposes to use a zero-truncated Poisson model to analyze human cancer tissues transplanted to mice when the positive counts of affected ducts is subject to right censoring. A Bayesian approach choosing a Gamma distribution as the prior is adopted. After implementing through complex computational procedures, this paper obtains the estimates of the coefficients and demonstrates model fitting through

the Deviance Information Criterion. The method is applied to investigate breast cancer progression using ductal carcinoma *in situ* data.

Liu, Hsu and Chen conduct a simulation study to evaluate gene selection models used in microarray gene expression data. Three predictive models that include data-reduction methods in the framework of a Cox proportional hazard model are evaluated. Log-rank test, prognostic index and deviance are the criteria for the evaluation. This paper may help medical researchers choose the appropriate model when conducting gene selection.

Das proposes to use a joint modeling technique to study the relationship between liver biochemical parameters and wine drinking. In this paper, the two joint models are the generalized linear log-normal model and the generalized linear gamma model. This unique analytic strategy in data analysis adopts the modern joint modeling technique that allows for interaction between two models.

The paper by Janiashvili, Jibladze, Matcharashvili and Topchishvili presents a method for comparing statistical characteristics of several 24-hour cardiovascular measurements (systolic blood pressure (SBP), diastolic blood pressure, pulse pressure, the mean arterial pressure and heart rate) on different groups discretized by SBP. A thorough discussion on the results is also presented.

These papers present several novel developments and applications of statistical methods to medical research which we hope readers will find useful in their own applications or methodological advancement.

Reference

[1]   Mayer, D. (2010) "Essential evidence-based medicine", Cambridge University Press. 2nd ed.