

The Many Facets of Statistical Modeling

**THE** INAUGURAL LECTURES are given by honored faculty members within the University who have obtained the rank of full professor. This event gives the honoree the opportunity to deliver a lecture to fellow faculty and other university guests concerning their work and research interests.

The context of the lecture itself typically includes a summary of the evolution and nature of the honoree's specialized field, highlights of some of the general issues of that particular field, and a description of how the honoree situates his/her work within their field.

UPM conducts this event to highlight and bring attention to the scholarly work that is being done by its distinguished faculty and to illustrate how the work contributes to mankind as a whole.

INAUGURAL LECTURE series

INAUGURAL LECTURE series

Prof. Dr. Noor Akma Ibrahim



Prof. Dr. Noor Akma Ibrahim

The Many Facets of Statistical Modeling



The Many  
Facets  
of Statistical Modeling

A portrait of Professor Dr. Noor Akma Ibrahim, a woman wearing a pink hijab and glasses. She is dressed in a maroon academic gown with a white collar and gold embroidery. The gown features a central panel with a repeating geometric pattern and wide sleeves with intricate gold floral and scrollwork designs. She is wearing a black mortarboard cap with a gold tassel. The background is a blurred bookshelf with several blue books.

**PROFESSOR DR. NOOR AKMA IBRAHIM**

# The Many Facets of Statistical Modeling

**PROFESSOR DR. NOOR AKMA IBRAHIM**

BSc and MSc (West. Mich. University) Ph.D (Universiti Putra Malaysia)

**22 NOVEMBER 2013**

Dewan Taklimat  
Universiti Putra Malaysia



**Universiti Putra Malaysia Press**

Serdang • 2013

<http://www.penerbit.upm.edu.my>

© **Universiti Putra Malaysia Press**

First Print 2013

All rights reserved. No part of this book may be reproduced in any form without permission in writing from the publisher, except by a reviewer who wishes to quote brief passages in a review written for inclusion in a magazine or newspaper.

UPM Press is a member of the Malaysian Book Publishers Association  
(MABOPA)

Membership No.: 9802

Typesetting : Juriah Mohamed  
Cover Design : Md Fairus Ahmad

*Design, layout and printed by*  
Penerbit Universiti Putra Malaysia  
43400 UPM Serdang  
Selangor Darul Ehsan  
Tel: 03-8946 8855 / 8854  
Fax: 03-8941 6172  
<http://www.penerbit.upm.edu.my>

## Contents

ABSTRACT	1
INTRODUCTION	3
SURVIVAL ANALYSIS	6
BAYESIAN INFERENCE	30
GENERALIZED LINEAR MODEL	33
CREDIBILITY MODEL	36
THE LANGAT WATER QUALITY INDEX	38
CONCLUSION	40
BIBLIOGRAPHY	40
BIOGRAPHY	51
ACKNOWLEDGMENT	53
LIST OF INAUGURAL LECTURES	55



## ABSTRACT

Statistics is the science of gaining and elevating insight from data. Data are pieces of information (often numerical but not always) gathered on people, objects or processes. The science of statistics involves all aspects of inquiry about data. Statistical modeling involves the finding of general laws from observed data, which amounts to extracting information from the data. Often the problem of main interest is to obtain a measure of both the complexity and the (useful) information in a set of data. Statistical modeling can be perceived also as a general framework for the application of statistical ideas. This presentation focuses on my experiences and endeavors in developing statistical models in order to capture the wide spectrum of data that arise from experiments, observations and other phenomena- producing data. Survival analysis is a collection of statistical models to explain the phenomena of survival data. The inherent property of survival data is the censored observation. There exists complexity in the censored mechanisms in which the usual survival models need to be modified or new survival models need to be developed to address this issue. One of the main objectives of survival analysis is to test the survival curves, and in the presence of partly interval-censored data goodness of fit tests have been developed for both parametric and nonparametric settings. Fundamental survival models do not include patients who are cured. In reality with the advancement of medical technology some patients are not susceptible to the disease (e.g. cancer) and this gives us the motivation to look into and consider models that will accommodate cure fraction for interval censored data with change point. When an event occurs and the death (example) of a patient can be due to several causes (risks), the survival model of interest is the competing risks model. A regression tree technique has been developed by using the subdistribution function of competing

risks to attain a better insight of a set of data. Bayesian is another approach that is fast gaining popularity in developing statistical models. Survival models with Bayesian approach considering several priors with right and interval-censoring are explored and developed. Regression Bayesian survival model with Jeffreys prior can be considered a frontier to Bayesian survival analysis. Simple and multiple linear regressions handle continuous response data with the assumption that the observations are independent. Substantial researches have been carried out to model scenarios of correlated data with binary and nominal response data. The models developed are based on Generalized Estimating Equation (GEE) for both semiparametric and nonparametric set-ups. Claim dependence model to include a third factor has been developed and its properties investigated. Due to certain limitations, the existing water quality index (WQI) measures which was based on experts opinions is very subjective in nature and does not provide an accurate picture of the water quality characteristics of a river. The subjectivity assumptions in developing WQI can be reduced by using statistical approaches. Moreover these statistical approaches can help to identify important parameters in determining the quality of a water body as well as the extent of their significance. Issues relating to this end this inaugural presentation.

## INTRODUCTION

*“Statistics...the most important science in the whole world: for upon it depends the practical application of every other science and of every art; the one science so essential to all political and social administration, all education, all organization based upon experience, for it only gives the results of our experience.”*

Florence Nightingale (1820-1910)

The world is complex and uncertainty exists. Indeed these complexities and uncertainties can be dealt with a variety of strategies that include scientific method and the discipline of statistics. Statistics helps to deal with uncertainty by quantifying it so that we can assess how reliable and likely the findings are. The scientific method helps to deal with complexity by reducing the systems to simpler components, defining and measuring quantities in the proper manner, and conducting experiments in which some conditions are held constant but varying others systematically. Beyond helping to quantify uncertainty and reliability, statistics provides insight that most people are unaware.

A well-defined study begins with a research question or hypothesis, devise a plan for collecting data, proceed to collect the data and analyze them, and then often make inference about how the findings generalize beyond the particular group being studied. Statistics concerns itself with all the phases of this process and therefore encompasses the scientific method.

A model is a representation for a particular purpose. Statistical models revolve around data. The intended use of a model should shape the appropriate form of the model and determines the sorts of data that can properly be used to build the model. Statistical

models can be used to describe, classify (or predict) and anticipate the consequences of interventions of the data. The appropriate form of a model depends on the purpose. For example, a model that diagnoses a patient as ill based on an observation of a high number of white blood cells can be sensible and useful. But that same model would give absurd predictions about intervention (Kaplan, 2009). The aim of a model is to capture aspects of a phenomenon that are relevant to inquiry and explain how the data could have come about as realization of a random experiment. These relevant aspects might include the genesis of randomness and the stochastic effects in the phenomenon under study. The defining characteristic of statistical models is their dependence on parameters and the incorporation of stochastic terms. The properties of the model and the properties of quantities derived from it must be studied in a long-run, average sense through expectations, variances and covariances. The parameters of the model that must be estimated from the data introduces the stochastic element in applying a statistical model. The model is thus not deterministic, it includes randomness whereby the parameters and related quantities derived from the model are likewise random. The properties of parameter estimators can often be described only in an asymptotic manner (the number of observations increases without bound). The process of estimating the parameters in a statistical model based on the data is known as fitting the model. The same model parameters can be estimated by different statistical principles, such as least squares method, maximum likelihood estimation, Bayes method and others. The parameter estimates obtained by different methods typically have different statistical properties (distribution, variance, bias and so on). The choice between these methods can be made based on the properties of the estimators. The distinguishing properties include computational ease, interpretive ease, bias, variance, mean squared error and consistency.

With the revolutionary advances in computer hardware and numerical algorithms over the last fifty years, complex statistical calculations can be performed. With appropriate software, any method is accessible in the sense of being able to produce a summary report on the computer. Nonetheless, the method is useful only when the user has a way to understand whether the method is appropriate for the situation, that is what the method is telling about the data and what the method is not capable of revealing. According to Richard Hamming (1915-1998) a computer scientist, the purpose of computing is insight, not numbers. Thus, to enrich data insight a solid understanding of the theory underlying a method is warranted.

This lecture showcases our contributions to the development of statistical models in tackling and capturing the multifaceted aspects of data in various situations. It will feature part of the vibrant research works being carried out over a decade while supervising several post-graduate students at UPM. In survival analysis, models have been developed particularly for partly interval-censored data parametrically and non-parametrically. The cure fraction models for both mixture cure model (MCM) and bounded cumulative hazard (BCH) were investigated and modified to accommodate interval-censored data with change point. To get a better insight of a breast cancer data, competing risks decision tree model based on the subdistribution function has been developed. Vigorous researches were conducted in the analyses of longitudinal data where the assumption of independence was not taken into consideration. The generalized estimating equation (GEE) was adopted in the modeling. Bayesian approach of implementing survival models was also considered. Non-informative prior in particular Jeffreys prior was considered in the formulation of survival regression model. Other works that will be discussed include the three-level

claim model and a water quality index for Langat river by using the established subindices which will execute a statistical perspective approach.

## **SURVIVAL ANALYSIS**

Survival analysis is one of the statistical methods which is important in the analysis of lifetime, particularly in medical and biological sciences. The outcome variable of interest is time recorded until an event occurs. In the context of medical and biological studies, the event of interest is often death, the onset of a disease or the disappearance of a disease's symptoms. The time to event of interest is called either survival time or failure time and the probability that a subject survives beyond a specified time is calculated by a basic quantity known as survival function.

The observation of survival time has two components that must be unambiguously defined: a beginning point and an endpoint that is reached when the event of interest occurs. One of the complications arises in survival analysis is when the observation is incomplete. Two mechanisms can lead to incomplete observation of time, and they are censoring and truncation. A truncated observation is due to the selection process. Different circumstances can produce many types of censoring. The most commonly encountered form of a censored observation is one where observation begins at the defined time and terminates before the outcome of interest is observed. Such observations are said to be right censored. Left censoring is encountered when the actual survival time is less than that observed. Interval censoring arises when the event of interest cannot be directly observed and it is only known to have occurred during a random interval of time.

The basic quantity employed to describe failure time phenomena is the survival function, the probability of an individual surviving beyond time  $t$ . It is defined as

$$S(t) = P(T > t)$$

where  $T$  is a non-negative random variable denoting failure time. If  $T$  is a continuous random variable, the survival function is the complement of a cumulative distribution function,

$$S(t) = 1 - F(T)$$

where  $F(t) = P(T \leq t)$ . The survival function is integral of the probability density function, that is

$$S(t) = P(T > t) = \int_t^{\infty} f(u) du$$

thus

$$f(t) = - \frac{dS(t)}{dt}$$

When  $T$  is a discrete random variable, a different technique is required. In addition to the survival function, the hazard function and the cumulative hazard function of  $T$  are also fundamental in survival analysis. Hazard function is also known as the conditional failure rate in reliability and can be defined as

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t | T \geq t)$$

If  $T$  is a continuous random variable, then

$$h(t) = \frac{f(t)}{S(t)} = \frac{d \ln[S(t)]}{dt}$$

A related quantity is the cumulative hazard function  $H(t)$ , defined by

$$H(t) = \int_0^t h(u) du = -\ln[S(t)]$$

It is easy to see that

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(u) du\right]$$

$S(t)$ ,  $h(t)$  or  $H(t)$  uniquely determines the distribution of  $T$ .

In an applied setting, the task of model selection is, to a large extent, based on the goals of the analysis and on the measurement scale of the outcome variable. There are many issues involved in the fitting, refining, evaluating and interpreting each of the models but the same basic modeling paradigm would be followed in each scenario.

A common feature for univariate survival data is right censoring. It occurs when the lifetime of an observation is only known to be greater than a known time. Censoring is assumed to be non-informative, that is, the censoring time  $C_i$  for individual  $i$  is statistically independent of the failure time  $T_i$ . For each observation, a bivariate random variable is observed  $(T_i^*, \delta_i)$ , where  $T_i^* = \min(T_i, \delta_i)$  and  $\delta_i = I(T_i < C_i)$  is a censoring indicator taking the value of 1 if  $T_i \leq C_i$  and 0 if  $C_i < T_i$ , right-censored. Given a simple random sample  $(t_i, \delta_i)$ ,  $i = 1, \dots, n$ , the likelihood function is given by

$$L = \prod f(t_i)^{\delta_i} S(t_i)^{1-\delta_i},$$

where  $S(t)$  is the survival function and  $f(t)$  is the corresponding probability density function.

### **Inference For Partly Interval-Censored Data**

Large body of the literature pertaining to censoring considers exact failure to be a special case of interval-censored. This gives us the notion to investigate more on the situation where they are seen as separate items better known as Partly interval-censored (PIC) data. The Partly interval-censored (PIC) data can occur in medical and health studies that are followed by periodic follow-ups. With PIC data, the failure times are exactly observed for some subjects but for the remaining subjects, the failure times are observed only to lie in an interval. An example of this kind of data is provided by the Framingham Heart Disease study whereby the time of the first emergence of angina pectoris in coronary heart disease patients was the event of interest. For some patients, the event time is recorded exactly, but for the remaining patients, time is recorded only to fall within the interval between two clinical examinations. There exist few researches that address the PIC such as Huang (1999), deriving the asymptotic properties of the nonparametric estimation for the distribution function with PIC data. Kim (2003) studied the maximum likelihood estimation in the presence of PIC data under the proportional hazards model. Zhao *et al.* (2008) developed a nonparametric test approach in the existence of the PIC data, which is based on the same idea of the generalized log-rank test for interval-censored data that was given in Sun *et al.* (2005).

Comparison of survival functions is one of the main objectives in survival studies. In most previously published research this comparison of survival distribution for two samples or more has been solved for cases with right-censored and interval-censored. We focused on the aspect of inferential comparison problem for survival functions in the presence of PIC, both parametrically and nonparametrically. We employed the multiple imputation method to reduce the interval censored data to exact data, which can be handled by using specified methods for exact data

Consider a schematic follow-ups for medical studies in which  $x_1, x_2, \dots, x_m$  are inspection times and the patient may be absent from the follow-ups with probability  $q'$  except at the first follow-up. Let  $T_i > 0$  be a random variable to denote the failure time of interest for  $i^{\text{th}}$  subject and  $n$  the number of subjects with failure times following a continuous distribution with density function  $f(t, \theta)$ , where  $\theta$  is a parameter vector. Assume that we observe the exact failure time for  $n_1$  subjects and interval-censored failure time for the remaining  $n_2$  subjects ; ( $n_2 = n - n_1$ ). By exact failure times we mean that any patient has the event of interest during the inspection times or that the patient's condition necessitates entering the hospital to be under examination, where the event of interest is recorded exactly. Also, by interval-censored failure times we mean that the event of interest occurs between two inspection time,  $(L_i, R_i)$  where  $L_i, R_i \in (x_1, x_2, \dots, x_m)$  and  $L_i < R_i$  with probability one. In addition if the event of interest happens for the patient before the first examination, then we have left-censored, i.e.  $t_i \in (0, L_i]$  and if the patient did not experience the event of interest at the last examination, we will have right-interval censored, i.e.  $t_i \in (R_i, \infty)$ . Also, we assume that censoring is independent of the examination time. For the  $i^{\text{th}}$  patient with interval-censoring define  $\delta_i = I(t_i \in (0, L_i)$  and  $\gamma_i = I(t_i \in (L_i, R_i])$ . Then the log likelihood function for  $\theta$  is

$$\begin{aligned} \ell = \log L(\theta) &= \sum_{i=1}^m \log f(t_i, \theta) + \sum_{i=m+1}^n \delta_i \log[1 - S(L_i, \theta)] \\ &+ \sum_{i=m+1}^n \gamma_i \log\left(\frac{S(L_i; \theta)}{S(R_i, \theta)} - 1\right) + \sum_{i=m+1}^n (1 - \delta_i) \log[S(R_i, \theta)] \end{aligned}$$

For the parametric case with  $T_i$  following the Weibull distribution, the hazard function has the form

$$h(t) = \left(\frac{a}{b}\right) \left(\frac{t_i}{b}\right)^{a-1}$$

where  $a$  and  $b$  are the shape and scale parameters respectively. Also assume that the PIC data are available. Under this assumption the log likelihood function is

$$\begin{aligned} L(a, b) &= \sum_{i=1}^m \log a - \log b + (a - 1)(\log t_i - \log b) - \left(\frac{t_i}{b}\right)^a \\ &+ \sum_{i=m+1}^n \delta_i \log[1 - \exp[-(\frac{L_i}{b})^a]] - \sum_{i=m}^n (1 - \delta_i) \left(\frac{R_i}{b}\right)^a \\ &+ \sum_{i=m+1}^n \gamma_i \log[\exp[(\frac{R_i}{b})^a - (\frac{L_i}{b})^a] - 1] \end{aligned}$$

Under Weibull model the maximum likelihood estimates of  $a$  and  $b$  are the solutions of partial derivatives of the log likelihood with respect to  $a$  and  $b$  and setting both to zeroes. To solve these simultaneous equations, we used Newton-Raphson method to obtain  $a$  and  $b$ .

The parametric tests considered were the score test and likelihood ratio test and these tests were extended to accommodate PIC data by using two methods. The first method was the generalized parametric test without multiple imputation technique (Direct Approach) and the second method was generalized parametric

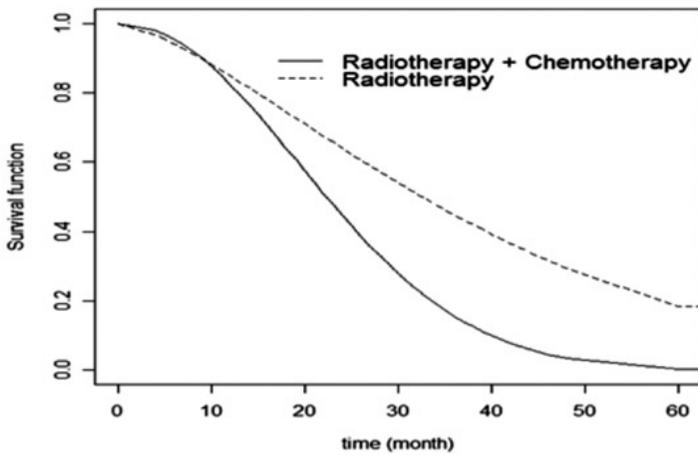
test via multiple imputation technique (Indirect Approach) (Azzah *et.al.*, 2012).

The aim is to test the hypothesis  $H_o : S_1(t) = \dots = S_p(t)$ , which is to determine whether the  $p$  treatments could have resulted from an identical failure time distribution. Let  $S_o(t)$  denotes the common survival function under  $H_o$  and  $\hat{S}_n(t)$  its parametric maximum likelihood estimate (PMLE) which is determined by estimating the parameter by using Newton-Raphson method.

Simulation studies were carried out to investigate and assess the performance of the estimators. The indirect approach is more efficient than the direct approach in most situation considered. For evaluating the performance of these two tests, with respect to the power, likelihood ratio test with indirect approach is better than the score test when the sample size is less than 50.

For real data application, we carried out some modification on the breast cancer study that was presented by Klein and Moeschberger (1997) so that it is applicable to apply the parametric test for PIC data. The target of this study is to compare two treatments with respect to their cosmetic effects. The data set are shown in Table 1. To test the cosmetic effect between two treatments, we used four tests: the score test with direct (STD) and indirect (STI) approaches and likelihood ratio test with direct (RTD) and indirect (RTI) approaches and the obtained values of the tests statistic equals to 23.775, 7.655, 15.085 and 7.262 respectively with p-values equals to 0.000, 0.022, 0.000 and 0.026 respectively. These results indicate that there is a significant difference between the treatments in terms of cosmetic effect. Figure 1 indicates that the patients in the (R+C) group develop breast retraction earlier than those in the R group.

In the nonparametric setting, the Turnbull self-consistency algorithm was modified in terms of PIC. The log-rank test was adopted with multiple imputations and evaluated. The proposed test behaves accordingly, in fact more accurate than Huang's test (2008).



**Figure 1** PMLE of survival function of time to cosmetic deterioration

**Table 1.** Time to cosmetic deterioration (in months) in breast cancer patients with two treatments

Radiotherapy (R)	Radiotherapy plus chemotherapy (R+C)
(0,7]; (0,8]; (0,5]; (4,11]; (5,12]; (5,11]; (6,10]; (7,16]; (7,14]; (11,15]; (11,18]; $\geq 15$ ; $\geq 17$ ; (17,25]; (17,25]; $\geq 18$ ; (19,35]; (18,26]; $\geq 22$ ; $\geq 24$ ; $\geq 24$ ; (25,37]; (26,40]; (27,34]; $\geq 32$ ; $\geq 33$ ; $\geq 34$ ; (36,44]; (36,48]; $\geq 36$ ; (37,44]; $\geq 37$ ; $\geq 37$ ; $\geq 37$ ; $\geq 38$ ; $\geq 40$ ; $\geq 45$ ; $\geq 46$ ; $\geq 37$ ; 12; 15; 17; 4; 18; 20; 21; 28; 30; 34; 22; 40; 42; 46; 36	(0,22]; (0,5]; (4,9]; (4,8]; (5,8]; (8,12]; (8,21]; (10,35] (10,17]; (11,13]; $\geq 11$ ; (11,17]; $\geq 11$ ; (11,20]; (12,20]; $\geq 13$ ; (13,39]; $\geq 13$ ; $\geq 13$ ; (14,17]; (14,19]; (15,22]; (16,24]; (16,20]; (16,24]; (16,60]; (17,27]; (17,23]; (18,25]; (18,24]; (19,32]; $\geq 21$ ; (22,32]; (17,26]; $\geq 23$ ; (24,31]; (24,30]; (30,34]; (30,36]; $\geq 31$ ; $\geq 32$ ; (32,40]; $\geq 34$ ; $\geq 34$ ; $\geq 35$ ; (35,39]; (44,48]; $\geq 48$ ; 4; 7; 10; 13; 15; 19; 22; 14; 18; 25; 29; 30; 32; 34; 40

## **Cure Fraction Models for Interval Censoring With a Change-Point**

The widely used model in survival analysis is the Cox (1972) proportional hazards model. This model is based on the assumption that every element in the population under study is susceptible to the adverse event of interest. But with the advances of medical treatment and health care, there are patients who are not susceptible to the occurrence of the event of interest. The proportion of such patients is considered as the cured fraction. The assumption of Cox is then violated and a more appropriate model is called for. Apart from that, in clinical trials or follow-up study, changes might occur at some unknown point that might need to be identified. Megan *et al.* (2012), developed a cure survival model that allows for change-point effects in covariates to investigate a potential change-point in the age of diagnosis of prostate cancer. However their work was limited to right-censored and confined to mixture cure model. Motivated by Megan's work, in the setting of interval-censored data we extended both the mixture cure model and the non-mixture cure model considering a change-point.

A change point problem is a problem in which changes at unknown points need to be identified and the locations of these changes need to be estimated. With this information, we could elucidate further information about the relationship between the covariate and the probability of survival. Change point problems occur frequently in survival studies. For example, data obtained from a group of preschool boys indicates that their weight/height ratio and their age have one functional relation before a certain age but that functional relation changes after this age (Gallant, 1977). As another example, cancer incidence rates remain relatively stable in younger people, but change drastically after a certain age (MacNeill and Mao, 1995).

### ***Mixture Model with Change-point***

Consider a group of patients entering a clinical trial. Let  $p$  be the proportion of patients cured on treatment and  $1-p$  be the proportion uncured. The survival function for the entire population of patients entering the clinical trial is given by

$$S(t) = p + (1-p)S_u(t),$$

where  $S_u(t)$  is the survival function for the uncured group. Suppose there are  $n$  patients entering a study. Let  $t_i$ ,  $i = 1, 2, \dots, n$  be the observed survival time for the patient. Let  $\delta_i$  be a censoring indicator defined such that  $\delta_i = 1$  if  $t_i$  is a failure and 0 if it is right-censored. Then the likelihood function is given by

$$L = \prod_{i=1}^n \{(1-p)f_u(t_i)\}^{\delta_i} \{p + (1-p)S_u(t_i)\}^{1-\delta_i}$$

where  $f_u(t_i)$  is the density function of uncured patients.

Consider a setting in which the event time  $T$  is known to have occurred within two time points  $(t_L, t_R)$ . Here,  $t_L$  is the latest examination time before the event, and  $t_R$  is the earliest examination time after the event, where  $t_{Ri} = \infty$  if subject  $i$  has not met the event before the last follow-up. Then, the observed data denoted by  $(t_{Li}, t_{Ri}, X_i)$ , can be written as

$$\begin{aligned} P(L_i \leq T_i \leq R_i) &= P(T_i \geq L_i) - P(T_i \geq R_i) \\ &= (1-p)[S_u(t_L | X) - S_u(t_R | X)] \end{aligned}$$

We can reformat the censoring indicator  $\delta_i$  as follows:  $\delta_i = I(t_R < \infty)$  for  $t_{Li} \leq t_i \leq t_{Ri}$ . Then, the likelihood function for the  $n$  observed interval-censored data is the following:

$$L = \prod_{i=1}^n [(1 - p[S_u(t_L | X) - S_u(t_R | X)])^{\delta_i} [p + (1 - p)S_u(t_L | X)]^{1 - \delta_i}]$$

Now, assume that the change point of the model depends on the covariate  $X$ , and that at this point or threshold the probability of cure or hazard function takes a sudden jump or fall. Suppose that the change point is  $\tau$ . Both  $\mu$  and  $p$  are thus dependent on the  $X$ , and if  $X \leq \tau$ , let  $p(X) = p_1$  and  $\mu(X) = \mu_1$ , while, if  $X > \tau$ ,  $p(X) = p_2$  and  $u(X) = u_2$ . Thus, by incorporating the unknown change point  $\tau$ , we can write

$$p(X) = p_1 I(X \leq \tau) + p_2 I(X > \tau) \text{ and} \\ \mu(X) = \mu_1 I(X \leq \tau) + \mu_2 I(X > \tau),$$

where the  $I(\cdot)$  indicator function taking the value 1 when the  $X \leq \tau$  and the value zero otherwise. Considering lognormal distribution, the unknown parameters are  $\theta = (p_1, p_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \tau)$ . The above likelihood then becomes:

$$L_n(\theta) = \left\{ \prod_{i=1}^n [(1 - p_1[S_{1u}(t_L, | X) - S_{1u}(t_R | X)])^{\delta_i} [p_1 + (1 - p_1)S_{1u}(t_L | X)]^{1 - \delta_i}]^{I(X_i \leq \tau)} \right. \\ \left. \times \left\{ [(1 - p_2)[S_{2u}(t_L, | X) - S_{2u}(t_R | X)]]^{\delta_i} [p_2 + (1 - p_2)S_{2u}(t_L | X)]^{1 - \delta_i} \right\}^{I(X_i > \tau)} \right.$$

### *Estimation Procedures*

Using the traditional likelihood approach, we will encounter two difficulties during the estimation process. Firstly,  $\tau$  is unknown and the likelihood function is not differentiable with respect to the change point parameter  $\tau$ , and consequently, standard Taylor series methods cannot be used. The second, is that the computation of the maximum likelihoods is complicated.

### *Smoothed Likelihood Approach*

To circumvent the critical problem of non-smoothing, a smoothed likelihood approach is proposed. The idea of this approach is to use a continuous and differentiable function to approximate the indicator functions  $I(X \leq \tau)$  and  $I(X > \tau)$ . Let  $K(u)$  be a continuous function such that  $K(u)$  is differentiable and non-decreasing over the real line, with  $\lim_{u \rightarrow -\infty} K(u) = 0$  and  $\lim_{u \rightarrow \infty} K(u) = 1$ . Define  $K(u) = K(u/h_n)$ , and  $h_n$  is a small positive constant that depends on the sample size. A useful special case of this class of function is the logistic function, where  $K(u) = \frac{\exp [u/h_n]}{1 + \exp [u/h_n]}$ . Based on this function, the smoothed likelihood for the observed data  $(\tilde{T}_i, \delta_i, X_i)$  is

$$L_n^*(\theta) = \prod_{i=1}^n \left\{ \left[ (1 - p_1) [S_{1u}(t_L | X) - S_{1u}(t_R | X)] \right]^{\delta_i} \right. \\ \left. \left[ p_1 + (1 - p_1) S_{1u}(t_L | X) \right]^{1 - \delta_i} \right\}^{K_n(\tau - X_i)} \\ \times \left\{ \left[ (1 - p_2) [S_{2u}(t_L | X) - S_{2u}(t_R | X)] \right]^{\delta_i} \right. \\ \left. \left[ p_2 + (1 - p_2) S_{2u}(t_L | X) \right]^{1 - \delta_i} \right\}^{1 - K_n(\tau - X_i)}$$

The second difficulty with the likelihood function is that the hazard function of  $T$  is no longer proportional if the cure fraction  $p$  is not equal to zero; thus, the simple form of the likelihood function cannot be obtained here (Kuk and Chen,1992). To solve this difficulty, we rewrite the likelihood using partially complete censored observations. Thus, the complete likelihood function can be changed to the modified likelihood function

$$L_n^*(\theta) = L_n \prod_{i=1}^n \left\{ [(1 - p_1)[S_{1u}(t_L, | X) - S_{1u}(t_R X)] \right]^{\delta_i} [(p_1)^{1-\eta_i} ((1 - p_1)S_{1u}(t_L | X)^{\eta_i})^{1-\delta_i}]^{K_n(\tau-X_i)} \times \left\{ [(1 - p_2)[S_{2u}(t_L, | X) - S_{2u}(t_R | X)] \right]^{\delta_i} [(p_2)^{1-\eta_i} + ((1 - p_2)S_{2u}(t_L | X)^{\eta_i})^{1-\delta_i}]^{1-K_n(\tau-X_i)}$$

where,  $\eta_i$  is not fully observed, i.e. , if  $\delta_i = 1$  then  $\eta_i = 1$ but if  $\delta_i = 0$ ,  $\eta_i$  is not observed and it can be one or zero. Therefore, to perform maximum likelihood estimation for the parameters  $\theta$ , we need to implement EM algorithm.

As usual simulation studies were conducted to investigate the performance of the proposed model. The simulation study conducted showed that based on the bias and mean square error values, the proposed estimation method performs well in the situations considered.

***Bounded Cumulative Hazard Model with Change-Point***

Chen *et al.* (1999) defined the bounded cumulative hazard (BCH) model as follows. Let  $N$  denote the number of carcinogenic cells that remain active to develop a cancer for  $i^{th}$  subject. Assume that  $N$  has a Poisson distribution with mean  $\theta$ . Let  $Z_j, j = 1, 2, \dots, N$  denote the random time for the  $j^{th}$  cancer cell to produce a detectable cancer mass, where,  $Z_j$  are assumed to be independently and identically distributed with  $F(t) = 1 - S(t)$ . Then the time to relapse of cancer can be defined by the random variable  $T = \min \{Z_j, j = 1, 2, \dots, N\}$ . The survival function for the population is given by

$$\begin{aligned} S_p(t) &= P[\text{No cancer by time } t] \\ &= P[N = 0] + P[Z_1 > t, Z_2 > t, \dots, Z_N > t, N \geq 1] \\ &= \exp(-\theta) + \sum_{N=1}^{\infty} (S(t))^N \left[ \frac{\exp(-\theta)^N}{N!} \right] \\ &= \exp(-\theta F(t)) = \pi^{F(t)} \end{aligned}$$

where  $\pi$  is the probability of cure which can be defined as  $\pi = \lim_{t \rightarrow \infty} S_p(t) \equiv P(N = 0) = \exp(-\theta)$ .

Let  $y_i$  denote the survival time for individual  $i$ , which might be right censored;  $y_i = \min(T_i, C_i)$  in which  $T_i = \min \{Z_{i1}, Z_{i2}, \dots, Z_{iN_i}\}$ , and  $C_i$  is right censored variable. Let  $\delta_i$  denote the censoring indicator, which equals 1 if  $y_i$  is failure time and 0 if it is right censored. Considering that censoring times are independent and non-informative, Chen *et al.* (1999), Mizoi (2004) and Weston and Thompson (2010) showed that the likelihood function for the model takes the form

$$L_i = \prod_{i=1}^n [-\log(\pi) f(y_i)]^{\delta_i} S(y_i)$$

We can further incorporate covariates  $X$  into both the cured probability and distribution function of uncured. Moreover, a parametric model can be specified for the survival time.

Suppose that the change-point model depends on  $X$ , and that at the change-point  $\tau$  the survival rate or cure probability takes a sudden jump or fall. Both  $\mu$  and  $\pi$  are depended on the  $X$ , then if  $X \leq \tau$ , write  $\pi(X) = p_1$  and  $\mu(X) = \mu_1$  while, if  $X > \tau$ ,  $\pi(X) = p_2$  and  $\mu(X) = \mu_2$ . In other words, we can write

$$\pi(X) = p_1 I(X \leq \tau) + p_2 I(X > \tau) \text{ and } \mu(X) = \mu_1 I(X \leq \tau) + \mu_2 I(X > \tau)$$

The complete data are  $(y_i, \delta_i, X_i)$  and the unknown parameters are defined by  $\theta = (p_1, p_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \tau)$ . Hence, the likelihood function under change-point  $\tau$  is defined as:

$$L_n = \prod_{i=1}^n \{[-\log(p_1)f_1(\theta, y)]^{\delta_i} p_1^{F_1(y)}\}^{I(X_i \leq \tau)} \\ \{[-\log(p_2)f_2(\theta, y)]^{\delta_i} p_2^{F_2(y)}\}^{I(X_i > \tau)}$$

With the classical likelihood approach, this likelihood function is not differentiable with respect to the change point  $\tau$ . Consequently, standard Taylor series methods cannot be used. A smoothed likelihood function is proposed.

### *Simulation Study*

The simulation study carried out 500 replications of sample sizes 250, 500, 1000, and 3000 for both models. Large sample sizes were needed to observe the asymptotic properties of smoothed parameters. Two simulation scenarios were considered. The first scenario used a Uniform (0, 1) random variable with a change-point at 0.5, while the second used a truncated Normal (1, 1, 0,

2) random variable with a change-point at 1. The event time was generated from the change point model with lognormal distribution, and the cure indicator was generated using uniform distribution to determine whether someone is cured.

The bias in the estimates for all the parameters reduced with increasing sample size for both normally and uniformly distributed covariate. Increasing sample size ensures that the sample characteristics get closer to the properties defined by the data generating model/process hence reducing bias. This observation demonstrates that the estimator of the parameters is statistically consistent. The Monte Carlo standard errors also reduced with increasing sample sizes across all parameters. Given the consistency of the estimator and the increased precision with increasing samples size, the root mean squares errors (MSE) also reduced with increasing sample size. In particular in the mixture model, the distribution of the covariate is relatively more accurate and precise to estimate the change point when the covariate is normally distributed than when it has a flat distribution.

## Competing Risk

Often in life-testing situation, failure of an individual can be identified as one or more of  $J$  ( $J \geq 2$ ) mutually exclusive, but possibly dependent cause of failure. In other words, each individual is subject to  $J$  distinct risks referred to as competing risks threatening his/her life. Occurrence of one event precludes observation of the other events on the same individual (it is assumed that patient can fail only from one cause). Associated with cause  $j$ , there is nonnegative absolutely continuous random variable  $X_j$  representing the lifetime of individual when no other potential risks are present. Suppose

the termination time of an individual is defined as the time to first failure. Thus, lifetime of an individual is given by  $T = \min \{X_1, \dots, X_j\}$ . The available information is usually given by the pair  $(T, \delta)$ , where  $\delta$  indicates the cause of failure, i.e.  $\delta = j$  if  $T = X_j$ . The competing risks concept can appropriately be applied to many areas of study such as industrial reliability analysis, market transaction analysis and clinical trial on paired organs.

### **Decision Tree for Competing Risks**

A huge amount of data has been rapidly accumulated, due to the fast development of computer technology. A new data analysis problem has arisen in such situation. Data mining is used to find important “knowledge” from large databases. Decision tree as one of many data mining techniques is a popular approach for segmentation, classification and prediction by applying a series of simple rules. It has the advantage that researchers can easily understand and explain the results, since it is expressed by a tree structured diagram as a final output. Decision trees automatically constructed from data have been used successfully in many real-world situations. Their effectiveness has been compared widely to other automated data exploration methods and to human experts. Decision tree can provide an important methodology in every data mining tools.

The landmark work of a decision tree is the Classification and Regression Trees (CART) methodology of Breiman *et al.* (1984), who introduced a tree methodology for univariate discrete or continuous response. A different approach is C4.5 proposed by Quinlan (1992).

There is now quite a lot of work dealing with decision tree especially in survival analysis (Gordon and Olshen 1985, Segal 1988, Davis and Andersen 1989, LeBlanc and Crowley 1992 and

1993, Segal 1995, Huang *et al.* 1998, Segal 1992, Zhang 1998, Su and Fan 2004, Gao *et al.* 2004), but there are no decision tree methods for competing risks survival data. Since analysis of competing risks survival data is complex due to the presence of more than one cause of failure, then it should be useful to develop such method.

In Ibrahim *et al.* (2008), we extended the decision tree for competing risks survival time data analysis by utilizing the advantage of proportional hazards model of subdistribution. Other related articles are in Noor Akma *et al.* (2009), Abdul Kudus *et al.* (2009), Ibrahim and Kudus (2008), Abdul Kudus *et al.* (2008) and Noor Akma and Abdul Kudus (2009).

In the application of competing risks tree, we used breast cancer data from Fyles *et al.* (2004). Between December 1992 and June 2000, 639 women 50 years of age or older who had undergone breast-conserving surgery for an invasive adenocarcinoma 5 cm or less in diameter (pathological stage  $T_1$  or  $T_2$ ) were randomly assigned to receive breast irradiation plus tamoxifen, RT+Tam, (319 women) or tamoxifen alone, Tam, (320 women). Participating centers included the Princess Margaret Hospital, the Women's College Campus of the Sunnybrook and Women's College Health Science Centre in Toronto, and the British Columbia Cancer Agency centers in Vancouver and Victoria. Table 2 contains the list of variables and their description.

**Table 2** Description of variable in the breast cancer study

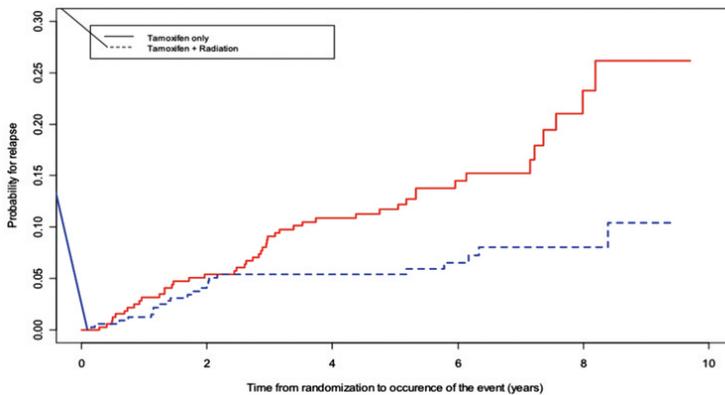
<b>Variable name</b>	<b>Description</b>
tx	Randomized treatment: 0=tamoxifen, 1=radiation+tamoxifen
<b>Variable assessed at the time of randomization</b>	
pathsiz	Size of the tumour (cm)
hist	Histology: 1 = ductal, 2 = lobular, 3 = medullary, 4 = mixed, 5 = other
hrlevel	Hormone receptor level: 0 = negative, 1=positive
hgb	Haemoglobin (g/l)
nodedis	Whether axillary node dissection was done, 0=Yes, 1=No
age	Age (years)
<b>Outcome variables</b>	
time	Time from randomization to event (relapse, second malignancy or death) or last follow up (years)
d	Status at last follow-up: 0=censored, 1=relapse, 2=malignancy, 3=death

The events that might occur in breast cancer study were relapse, second malignancy and death. The patient's survival time was the time length between the date of randomization and the occurrence of one event or last follow-up date.

Since the goal of regression tree is to partition patients into groups on the basis of similarity of their responses to treatment, we constructed a separate regression tree for each treatments (tamoxifen alone and tamoxifen plus radiation). The partitioning is based on baseline characteristics such as patient demographics and clinical measurements. The final tree structure provides treatment effect

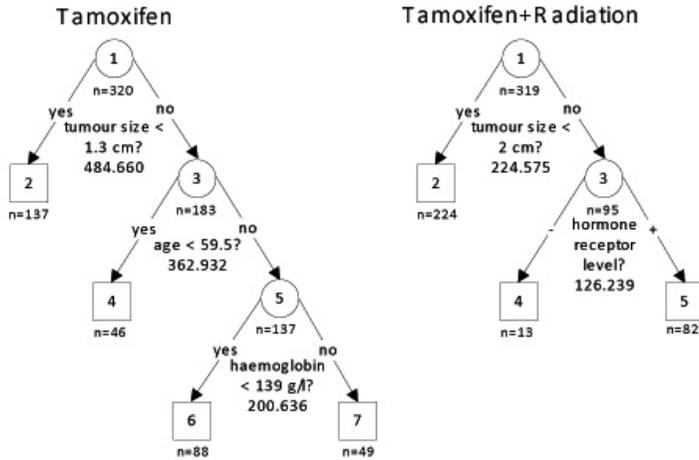
within each group of patients. The question to be answered by this type of analysis is – for whom does the treatment work best?

The cumulative probability for relapse by time  $t$  is shown in Figure 2. Here we compare the probability for two types of treatment. The patients with tamoxifen plus radiation have less probability to relapse compared to those with tamoxifen alone as expected. It showed the advantage of radiation in reducing the occurrence of relapse.



**Figure 2** Cumulative Probability for relapse for two types of treatment

The exploration was further employed to find group of patients for each treatment by using decision tree. With respect to probability for relapse, we obtained four groups of patients which were treated by tamoxifen alone, and three groups of patients which were treated by tamoxifen plus radiation (Figure 3).



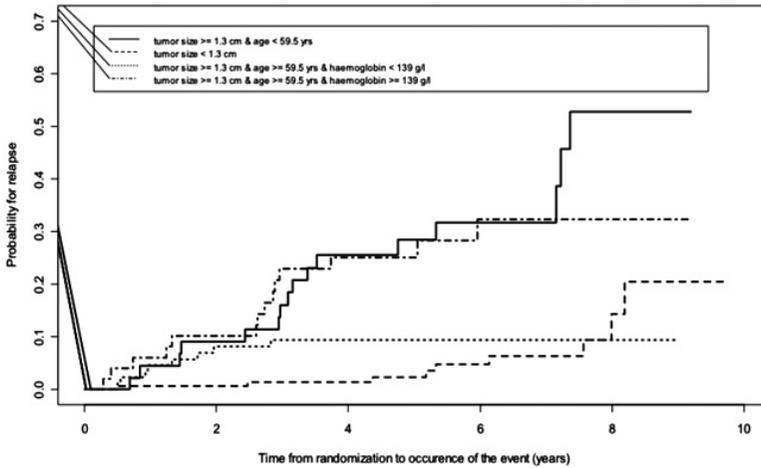
**Figure 3** Decision tree for probability for relapse

The description of four groups of patient which was treated by tamoxifen alone is:

1. Node 2: tumour size < 1.3 cm
2. Node 4: tumour size  $\geq$  1.3 cm and age < 59.5 years
3. Node 6: tumour size  $\geq$  1.3 cm and age  $\geq$  59.5 years and haemoglobin < 139 g/l
4. Node 7: tumour size  $\geq$  1.3 cm and age  $\geq$  59.5 years and haemoglobin  $\geq$  139 g/l

This group formation reveals that node 2 has the lowest probability to relapse up to about 8 years follow-up, whereas node 4 has the highest probability (Figure 4).

## The Many Facets of Statistical Modeling

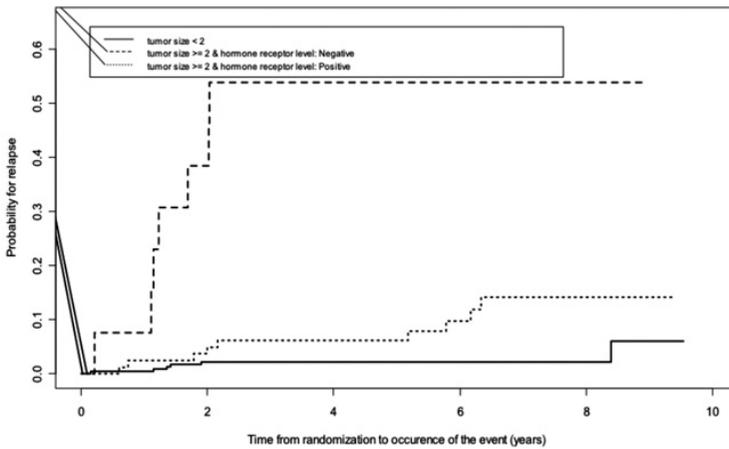


**Figure 4** Cumulative Probability for relapse for groups resulted by decision tree for patient with tamoxifen alone

For patients with tamoxifen plus radiation, there are three groups resulted by decision tree, namely:

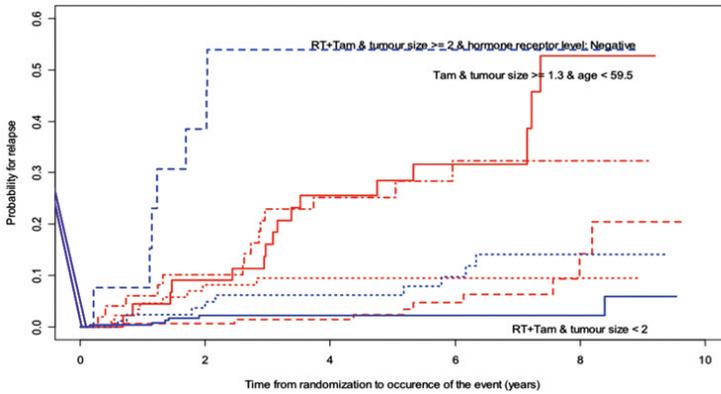
1. Node 2: tumour size  $< 2$  cm
2. Node 4: tumour size  $\geq 2$  cm and hormone receptor level negative
3. Node 5: tumour size  $\geq 2$  cm and hormone receptor level positive

Women with tamoxifen plus radiation whose tumour size less than 2 cm have the lowest probability to relapse, whereas the highest probability is for those whose tumour size  $\geq 2$  cm and negative hormone receptor level (Figure 5).



**Figure 5** Cumulative Probability for relapse for groups resulted by decision tree for patient with tamoxifen plus radiation

Overall comparison for both treatments reveals that the poorest and best prognosis are from tamoxifen plus radiation treatment group. We found that tamoxifen plus radiation is not effective for those women whose tumour size greater than 2 cm and negative hormone receptor level. This group is more likely to relapse compared to the others. On the other hand, patients with tamoxifen plus radiation and tumor size less than 2 cm have the best prognosis, because they are less likely to relapse (Figure 6).



**Figure 6** Cumulative Probability for relapse for all groups resulted by decision tree

The analysis for two other events (second malignancy and death) showed different results. This showed that patients give different responses to the treatment.

This competing risks tree method based on proportional hazards of subdistribution model intends to provide an exploratory data analysis for competing risks survival data, and it is complimentary rather than competitive to those parametric or semi-parametric methods. The application on breast cancer data showed that the method could find groups of data which had similar response to treatment. Simulation results showed that the proposed method performs well for prognostic classification. In all the simulations, high portion of the data structures can be correctly identified.

## BAYESIAN INFERENCE

Bayesian estimation approach has recently become a generally acceptable method in estimating parameters. Previously, the

Bayesian approach was discouraging due to the necessity of numerical integration. As a result of the radical change in the computer intensive sampling methods of estimation, the Bayesian method is now vigorously pursued by researchers for its comprehensive approach to the estimation of complex models. Bayesian inference is an approach that employs the Bayes' rule in order to update the probability estimate of a hypothesis taking into consideration new evidences as they become available. Bayesian updating is one of the essential techniques used in modern statistics, more importantly in mathematical statistics. The Bayes approach makes use of our prior beliefs of the parameters which is referred to as Prior distribution. The inference is based on the posterior estimate which is simply the combination of ones prior knowledge and the availability of data (the likelihood).

Non-informative prior is one of the categories of the prior distribution. It refers to a situation where there is very limited knowledge or information prior to the researcher. With non-informative prior there is little or no influential information that is added to the actual data available. What this means is that we have an occurrence of a set of parameter values in which the statistician believes that the choice of a parameter is equally likely. Jeffreys prior and extension of Jeffreys prior are used to avoid any hyper parameter specification. Both are invariant under reparametrization, because of the relation to Fisher information, when we have large information, we minimize the influence of the prior such that it is non-informative as possible. Jeffreys prior and extension of Jeffreys are very useful for data that do not have any prior information available and give better result in many cases than classical estimation.

## Bayesian Estimation Under Non-Informative Prior

There is a huge body of literature on Bayesian survival models. Bayesian with non-informative priors and right censoring (without consideration of covariates) can be seen in Al Omari *et al.* (2010), (2012) and Al Omari *et al.* (2011). We developed Bayesian models to incorporate covariates by employing Jeffreys and extension of Jeffreys as priors. Under the regression framework the extension of Jeffreys was modified to incorporate the covariates. The Weibull distribution was considered as the distribution of the time to event due to its appealing features that includes its ability to provide reasonably accurate analysis with extremely small samples. With the failure time  $(t_1, \dots, t_n)$  following Weibull distribution the proposed extension Jeffreys prior is

$$g(\beta', \theta) = k(\det[I(\beta', \theta)])^{\log(\sum_{j=0}^n \beta_j)},$$

where  $k$  is a constant and  $I(\beta', \theta)$  is the Fisher information matrix with covariates and an unknown shape parameter. The joint probability density function with right censoring is

$$H(\beta', \theta | t) = g(\beta', \theta) \prod_{i=1}^n [f(t_i; \beta', \theta)]^{\delta_i} [S(t_i; \beta', \theta)]^{1-\delta_i}$$

where  $f(\cdot)$  and  $S(\cdot)$  are the probability function and survival function respectively with censoring indicator  $\delta_i$ .

The marginal probability density function is

$$P(\beta', \theta | t) = \int_0^\infty \dots \int_0^\infty g(\beta', \theta) \prod_{i=1}^n [f(t_i; \beta', \theta)]^{\delta_i} [S(t_i; \beta', \theta)]^{1-\delta_i} d\beta' d\theta$$

The posterior probability density function is

$$\Pi(\beta', \theta | t) = \frac{H(\beta', \theta | t)}{P(\beta', \theta | t)}$$

The Gauss quadrature rule can be used to estimate each parameter with a chosen loss function.

The performance of Bayesian estimators was assessed and compared with its maximum likelihood counterpart via simulation study under several conditions. The results indicated that under certain situations modified Jeffreys emerges as a better method compared to the others. All methods produced estimators that behaved in the appropriate manner in terms of consistency. Models encompassing interval-censored data have also been developed based on Bayesian approach, see Chris *et al.* (2013). Other related works with Bayesian can be seen in Guure and Ibrahim (2013), Chris and Noor Akma (2012,2013), Chris *et al.* (2012a), (2012b).

## GENERALIZED LINEAR MODEL

Generalized linear model (GLM) is essentially a unified method of analyzing certain types of data situations. It is based on the exponential family of probability distributions which includes normal, binomial, Poisson, gamma, inverse Gaussian, geometric, and for a given auxiliary parameter, the negative binomial. The binomial models themselves include logit, probit, and log-log, to name a few. One may use GLM to model ordinary least squares (OLS) regression as well as logistic, probit and Poisson regression models. The ability to compare parameter estimates, standard errors and summary statistics between models gives the researcher a powerful means to obtain an optimal model for a given dataset. However, being likelihood based, GLMs assume that individual rows in the data are independent from one another. In the case of longitudinal data, this assumption may fail where the data

are correlated (Hardin and Hilbe, 2002). Hence, we proposed generalized estimating equation (GEE) with smoothing spline to capture the aspect of correlation in the data.

### **GEE-Smoothing Spline Model**

It is very common in economics, epidemiology or clinical trials to make a study in which subjects are followed over time or measured on several occasions to collect response variables. This type of study is commonly known as longitudinal study. The characteristic of these data is that they are no longer independent, in which there is correlation among within subject measurements. Another characteristic is that the variances usually are not homogeneous. Thus methods in the class of generalized linear model (GLM) are no longer valid for these data, since GLM assumes that observations are independent. Some developments have been proposed to analyze such data, that can be classified into three types of model, marginal model, subject specific effect, and transition model (Davis, 2002). In the class of marginal model, Liang and Zeger (1986) and Zeger and Liang (1986) extended *quasi-likelihood estimation* of Wedderburn (1974) by introducing “working correlation” to accommodate within subject correlation, which is called generalized estimating equation (GEE). GEE yields consistent estimates of the regression coefficients and their variances even though there is misspecification of the working correlation structure, provided the mean function is correctly specified.

GEE is part of the class of parametric estimation, in which the model can be stated in a linear function and the function is known. Very often the effect of the covariate cannot be specified in the specific function. Nonparametric regression can accommodate this problem by relaxing relationship between covariate and response. In nonparametric regression, we assume that the

effect of the covariate follows an unknown function without specific term, that is it is just a smooth function. To date there are several methods in nonparametric regression, for example: local polynomial kernel regression, penalized splines regression, and smoothing splines. Green and Silverman (1994) gave a simple algorithm for nonparametric regression using cubic spline by penalized least square estimation. They also gave nonparametric and semiparametric methods for independent observations for class of generalized linear models. We proposed GEE-smoothing spline to analyze longitudinal data and study the properties of the estimator such as the bias, consistency and efficiency. We used natural cubic spline and combine this with GEE of Liang & Zeger's in the estimation.

From the simulation, we can conclude that GEE-smoothing spline has better properties than GEE-local polynomial kernel proposed by Lin & Carroll (2000). The pointwise estimates of GEE-smoothing are consistent, even if we use incorrect correlation structure. The convergency rates of consistency for independent data (no correlation), moderate correlation, and high correlation are the same. If data are correlated, ignoring this correlation in the model, will give the most inefficient estimate. Taking into account the dependency into the model is better than ignoring it, even with incorrect correlation structure. If data are independent, the efficiency of using correct or incorrect correlation structures is almost similar. Hence, since in true situation the correlation is unknown, then it is better to assume the data are correlated rather than to assume data are independent (Suliadi *et al.*, 2010a). We extended this with semiparametric estimation (Suliadi *et al.*, 2010b). Comparison of some smoothing parameter selection methods can be seen in Suliadi *et al.* (2009). We also have developed an algorithm on how to generate correlated discrete ordinal data using R and

SAS IML (see Noor Akma and Suliadi, 2011). Another extension of semiparametric estimation with profile algorithm for longitudinal binary data is in Suliadi *et al.*, (2013).

## CREDIBILITY MODEL

One of the basic problems in presenting various insurance policy is to determine their premiums. If we have observations of past claims for a group of contracts then we might be able to predict the next period premium. Thus, for insurance risks, the determination of their premiums must strongly reflect the features of those risks. In insurance premium determination, it is a familiar practice to group individual risks to ensure homogeneity in reaching a fair and equitable premium across the individuals so that the risks within each group are as homogeneous as possible in terms of certain observable risk characteristics. That is a rating group is viewed as homogeneous with respect to the underwriting characteristics. However, not all risks in the group are truly homogeneous. A collective premium also called the 'manual premium' is then calculated and charged for this group. The collective premium is designed to reflect the expected experience of the entire rating class and implicitly assumes that the risks are homogeneous. In general credibility theory is a set of quantitative tools which allows an insurer to perform prospective experience rating (adjust future premiums based on past experience) on a risk or group of risks (Klugman *et al*, 2004). Based on the experience and the collective premium, the credibility theory determines the credibility premium by the following credibility form:

Credibility premium =  $Z \times (\text{experience}) + (1-Z)(\text{collective premium})$ ,

where  $Z$  a value between 0 and 1, is called the ‘credibility factor’ and needs to be chosen.

There have been a variety of insurance pricing methodologies to serve this purpose, among which is the credibility ratemaking. This is one of the most important techniques used in general insurance pricing. Under credibility techniques, one separately and adaptively determines the premiums for each contract in a heterogeneous portfolio, by effectively combining the policyholder’s claim experience and the portfolio’s particular risk features. In credibility models there are so called structure parameters that must be estimated before the calculation of the credibility estimators themselves.

In the usual credibility model, observations are made of a risk or group of risks selected from a population and claims are assumed to be independent between different risks. However, there are some problems in practical applications and it may be violated in some situations. Some credibility models typically allow for one source of claim dependence only, that is across time for an individual insured risk or group of homogeneous insured risks. There exist some other credibility models in the literature which have been developed on two-level common effects model that allows for two possible sources of dependence: across time for the same individual and that between individual risks. We established the notion of modeling claim dependence in credibility models with dependence induced by three-level common effects that allows for three possible sources of dependence: the dependence among portfolio risks, dependence of the individual risks and the dependence of experience for a particular individual risk over time. The properties of the model were investigated and the structural parameters estimated (Mahdi *et al.*, 2013 and 2011). The unbiased estimation of the structural parameters were derived (Mahdi *et al.*, 2012).

## THE LANGAT RIVER WATER QUALITY INDEX

Water quality is generally described according to biological, chemical and physical properties (Bharti and Katyal, 2011). Based on these properties, the quality of water can be expressed via a numerical index that is Water Quality Index (WQI) by combining measurements of selected water quality variables. The selected water quality variables were identified with respective weights and the determining processes were based on personal evaluation, namely, opinion gathering techniques. The weights assigned to the selected variables were based on the relative importance given by the experts. This weights determination technique is used by other researchers including Malaysian Department of Environment (DOE). The selected variables, together with respective weights are applied to calculate water quality index in all rivers in Malaysia. Due to varying characteristics for each river, the weights for water quality variables may be different for separate rivers. Therefore it is clear that the existing weights of the selected variables, as per DOE are subjective in nature and no detailed studies have been done to determine the weights objectively.

Langat River is one of the most important raw water resources for drinking water, recreation, industry, fishery and agriculture. The river which is situated in the state of Selangor, Peninsular Malaysia with a total catchment area of approximately 1,815km<sup>2</sup> is chosen for the pollution prevention improvement programme, introduced by the Malaysian Department of Environment (DOE) from 2001 to improve the condition of polluted rivers in Malaysia. Langat River as a tropical catchment area is experiencing rapid urbanization (Amini *et al.* 2009) and the gain in size of urbanised area was also reported by Jaafar *et al.* (2009). The increasing of developing areas within the river basin increases pollution loading into the Langat

River. Thus, surface water pollution is identified as the major problem affecting the Langat River Basin in Malaysia and major sources of surface water quality variations in Langat River come from industrial effluents, wastewater treatment plants, domestic and commercial areas (Juahir *et al.* 2010). It is also obvious that the Langat River ecosystem is under stress from the discharge of effluents particularly domestic sewage (Lee *et al.* 2006). Due to rapid urbanization and changes from undeveloped to developed area, Langat River experienced changes of pollutants with respect to space and time and we are unable to capture these drastic changes if WQI is not revised. Therefore, the revision of the river Water Quality Index (WQI) is needed as recommended by Juahir *et al.* (2010). Thus we undertake the task of determining the weights and subsequently formulate a new WQI by using the approach of multivariate and Bayesian. In Mohd Ali *et al.* (2013 a, and c), the relative importance of six water quality variables (dissolved oxygen, (DO); biochemical oxygen demand, (BOD); chemical oxygen demand, (COD); suspended solid, (SS); potential of hydrogen, (pH); ammonia, (AN)) were investigated by using cluster analysis followed by discriminant analysis. These variables were considered in the WQI calculation by DOE and were established in a formula based on expert opinions which is very subjective in nature that do not take into consideration the interrelationships between the variables. The results in the relative ranking were different and we believe the ranking from the statistical approach can be used as a guide to investigate other influential available variables in the water quality of Langat River. And in Mohd Ali *et al.* (2013 b), in the development of a new Bayesian S-W (Stock-Watson) index, the Bayesian model comparison criterion was used to choose the best coincident-index model.

## CONCLUSION

The role and the importance of statistical modeling are obvious as we need modeling for the purpose of statistical inference, describing phenomena, prediction and decision making. With the rapid progression of technology, challenges in statistical modeling are forever growing. Moreover, data are becoming more complicated and sophisticated and these entail a more creative and innovative statistical techniques for the statisticians to explore in order to capture the many facets of the data. To procure virtuous statistical models and enhance model performance the statisticians should strongly embrace three things (Abdul Aziz Jemain, 2011). First is to reinforce the mathematical statistics knowledge as this will help in the assimilation of the theory and application, enhance computational ability and strengthen problems formulation. Without doubt the second requirement is to have a profound computational skill. Third is the need to have the ability to comprehend strongly the fundamentals in order to translate and provide reasonable data interpretation and inference.

## BIBLIOGRAPHY

- Abdul Aziz Jemain (2011). *Memperluas Jelajahan Memperkaya Khazanah Ilmu Statistik*. Syarahan perdana jawatan Profesor, Universiti Kebangsaan Malaysia, Penerbit UKM, Bangi.
- Abdul Kudus, Noor Akma Ibrahim, Isa Daud & Mohd. Rizam Abu Bakar (2008). Hybrid Model for Subdistribution of Competing Risks. *Matematika*, Special Edition, Part II: 449-457.
- Abdul Kudus, Noor Akma Ibrahim, Mohd. Rizam Abu Bakar & Isa Daud (2009). Competing risk data generation based on subdistribution function. *International Journal of Applied Mathematics and Physics*, 1(1) Jan-June: 1-9.

- Ali, Z. M., Ibrahim, N. A., Mengersen, K., Shitan, M., & Juahir, H. (2013a). New relative importance of water quality variables in Langat River. *International Journal of Chemical and Environmental Engineering*, 4(1), 25-31.
- Ali, Z. M., Ibrahim, N. A., Mengersen, K., Shitan, M., & Juahir, H. (2013b). The selection of the Bayesian Coincident-index Models using Model Comparison Criterion with application in Langat River water quality data. *AIP Conf. Proc.* 1522, 1283(2013); doi: 10.1063/1.4801278.
- Ali, Z. M., Ibrahim, N. A., Mengersen, K., Shitan, M., & Juahir, H. (2013c). The Langat River Water Quality Index Based on Principal Component Analysis. *AIP Conf. Proc.* 1522, 1322(2013); doi: 10.1063/1.4801283
- Al Omari Mohammed Ahmed, Noor Akma Ibrahim, Mohd. Bakri Adam & Jayanthi Arasan (2012). Bayesian survival and hazard estimate for Weibull Censored Time distribution. *Journal of Applied Sciences*, 12,1313-1317.
- Al Omari Mohammed Ahmed, Noor Akma Ibrahim, Jayanthi Arasan & Mohd. Bakri Adam (2011). Extension of Jeffreys's Prior Estimate for Weibull Censored Data using Lindley's Approximation. *Australian Journal of Basic and Applied Sciences*, 5(12),884-889.
- Al Omari Mohamed Ahmed & Noor Akma Ibrahim (2011). Bayesian survival Estimator for Weibull Distribution with Censored Data. *Journal of Applied Sciences*, 11(2),393-396.
- Al Omari Mohammed Ahmed, Hadeel Salim Al-Kutubi & Noor Akma Ibrahim (2010). Comparison of the Bayesian and Maximum Likelihood estimation for Weibull Distribution. *Journal of Mathematics and Statistics*, 6(2), 100-104.
- Amini, A., Mohammad Ali, T, Ghazali, A.H., & Kim Huat, T. (2009). Adjustment of peak streamflows of a tropical river for urbanization. *American Journal of Environmental Sciences*, 5(3), 285-294.
- Azzah Mohamed Al Harpy, Noor Akma Ibrahim & Jayanthi Arasan. (2013). Parametric tests or partly interval-censored failure time data under Weibull Distribution via Multiple Imputation. *Journal of Applied Sciences*, 13(4),621-626.

- Bader Ahmad Aljawdi, Mohd. Rizam Abu Bakar & Noor Akma Ibrahim. (2012). Turnbull versus Kaplan-Meier estimation of cure rate estimation using interval-censored data. *Pertanika J. Sci & Technology*, 20(2), 243-255.
- Bader Ahmad Aljawdi, Mohd. Rizam Abu Bakar & Noor Akma Ibrahim. (2012). Nonparametric versus parametric estimation of the cure fraction using interval-censored data. *Comm, in Statistics-Theory and Methods*, 41(Oct),4251-4275.
- Bader Ahmad Aljawdi, Mohd. Rizam Abu Bakar & Noor Akma Ibrahim (2011). An analytical approach on parametric estimation of cure fraction based on Weibull distribution using interval-censored data. *Australian Journal of of Basic and Applied Sciences*, 5(3), 546-554.
- Bader Ahmad Aljawdi, Mohd. Rizam Abu Bakar, Noor Akma Ibrahim & Habshah Midi (2011). Parametric estimation of the immunes proportion based on BCH model and exponential distribution using left censored data. *Journal of Applied Sciences*, 11(15), 2861-2865.
- Berkson, J., Gage, R., Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47(1952), 501–515.
- Breiman, L, Friedman, J, Olshen, R., & Stone, C. (1984). *Classification and regression trees*. New York: Chapman and Hall.
- Bharti, N., & Katyal, D. (2011). Water quality indices used for surface water vulnerability assessment. *International Journal of Environmental Sciences*, 2(1),154-173.
- Boag, J.W. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society Series B*, 11, 15-44.
- Chen, M.-H., Ibrahim, J.G., & Sinha, D., (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, 94, 909–919.
- Chris Bambeby Guure, Noor Akma Ibrahim & Mohd. Bakri Adam (2013). Bayesian Inference of the Weibull Model based on Interval-Censored Survival Data. *Computational and Mathematical Methods in Medicine*, 2013, Article ID 849520, doi:10.1155/2013/849520.

- Chris Bambey Guure & Noor Akma Ibrahim (2013). Generalised Bayesian Non-Informative Prior Estimation of Weibull Parameter with Interval Censoring. *ScienceAsia*, 39S: 75-79.
- Chris Bambey Guure, Noor Akma Ibrahim & Al Omari Mohammed Ahmed (2012). Bayesian estimation of Two-Parameter Weibull Distribution using Extension of Jeffreys' Prior information with Three Loss Functions. *Mathematical Problems in Engineering*, Vol 2012,doi:10.1155/2012/589640.
- Chris Bambey Guure & Noor Akma Ibrahim (2012). Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data. *Mathematical Problems in Engineering*, Vol 2012,doi:10.1155/2012/329489.
- Chris Bambey Guure, Noor Akma Ibrahim & Mohd Bakri Adam (2012). On partly censored data with the Weibull distribution. *ARP Journal of Engineering and Applied Sciences*, 7(10), 13291334.
- Collett, D. (1994). *Modelling survival data in medical research*. London: Chapman and Hall.
- Jaafar, O., S.A. Sharifah Mastura & Mohd Sood, A. (2009). Land use and deforestation modelling of river catchments in Klang Valley, Malaysia. *Sains Malaysiana*, 38(5), 655-664.
- Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society: Series B*, 34(5),187-220.
- Davis, Charles S. (2002). *Statistical methods for the analysis of repeated measurements*. New York. USA: Springer-Verlag.
- Davis, R. & Anderson, J. (1989). Exponential survival trees. *Statistics in Medicine*, 8, 947-962.
- Fauzia Ali Taweab, Noor Akma Ibrahim & Bader Ahmad Aljawdi (2012). Estimation of cure fraction for lognormal right censored data with covariates. *International Journal of Modern Physics: Conference Series*, 9, 308-315.
- Fine, J. P. & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, 94, 496-509.

- Fyles, A. W., McCready, D. R., Manchul, L. A., Trudeau, M. E., Merante, P., Pintilie, M., Weir, L. M. & Olivotto, I. A. (2004). Tamoxifen with or without breast irradiation in women 50 years of age or older with early breast cancer. *New England Journal of Medicine*, 351, 963-970.
- Gallant, A. R. (1977). Testing a nonlinear regression specification: A non regular case. *J. Amer. Statist. Assoc.*, 72, 523–530.
- Gao, F., Manatunga, A. K., & Chen, S. (2004). Identification of prognostic factors with multivariate survival data. *Computational Statistics and Data Analysis*, 45, 813-824.
- Gordon, L. & Olshen, R. (1985). Tree-structured survival analysis. *Cancer Treatment Reports* 69, 1065-1069.
- Green, P. J., & B. W. Silverman. (1994). *Nonparametric regression and generalized linear models*. A roughness penalty approach. New York, USA: Chapman & Hall/CRC.
- Guure, C. B., & Ibrahim, N. A. (2013). Methods for estimating the 2-parameter Weibull Distribution with Type-I Censored Data. *Research Journal of Applied Sciences, Engineering and Technology*, 5, 689-694.
- Hardin, James W., & Joseph M Hilbe. (2003). *Generalized estimating equations*. Washington, DC. USA: Chapman & Hall/CRC.
- Huang, X., Chen, S. & Soong, S. (1998). Piecewise exponential survival trees with time-dependent covariates. *Biometrics*, 54, 1420-1433.
- Huang, J. (1999), Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statist. Sinica*. 9, 501–519.
- Ibrahim, J.G., Chen, M.-H., & Sinha, D.(2001). *Bayesian Survival Analysis*. New York: Springer.
- Ibrahim, N. A., & Kudus, A. (2008). On the data mining concepts and techniques with application in business. *Journal of Quantitative Methods*. 4(2), 85- 97.
- Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data*(2nd Ed). New York: Wiley.
- Kaplan, D. T.(2013). *Statistical modeling, a fresh approach*. 1<sup>st</sup> Ed. (Amazon).

- Kim, J. S. (2003), Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65, 489–502.
- Kim Y., & Jhun M.(2008). Cure rate model with interval censored data. *Statist. Med.*, 27, 3-14.
- Klein, J. B., & Moeschberger, M. L. (1997), *Survival analysis techniques for censored and truncated data* (2nd Ed.). New York: Springer..
- Klugman, S., Panjer, H., & Willmot, G. (2004). *Loss models: from data to decisions*. New York: John Wiley & Sons, Inc.
- Kuk, A., & Chen, C. (1992). A mixture model combining logistic regression with proportional hazards regression, *Biometrika*, 79, 531–541.
- Lambert, P.C., Thompson, J.R., Weston, C.L., (2007). Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, 8(3), 576-594.
- LeBlanc, M., & Crowley, J. (1992). Relative risk trees for censored survival data. *Biometrics* 48, 411-425.
- LeBlanc, M., & Crowley, J. (1993). Survival trees by goodness of split. *Journal of the American Statistical Association*, 88, 457-467.
- Lee, Y.H., Abdullah, M.P., Chai, S.Y., Mokhtar, M., & Ahmad, R. (2006). Development of possible indicators for sewage pollution for the assessment of Langat River ecosystem health. *Malaysia Journal of Analytical Sciences*, 10(1), 15-26.
- Leisch, Friedrich, Andreas Weingessel & Kurt Hornik. (1998). On the Generation of Artificial Correlated Binary Data. *Working Paper*. No 13. SFB. Adaptive Information System and Modeling in Economics and Management Science. Vienna University of Economics and Business Administration. Austria.
- Liang, K. Y., & S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- Lin, Xihong & Raymond J. Carroll.(2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95, 520-534.

- Lin, Xihong and Raymond J. Carroll. (2001). Semiparametric Regression for Clustered Data Using Generalized Estimating Equations. *Journal of the American Statistical Association*. 96, 1045-1056.
- Ma, Y. (2011). Testing change-point in logistic models with covariate measurement error. *Journal of Statistical Research*, 45, 131-138.
- MacNeill, I. B., & Mao, Y. (1995) Change-point analysis for mortality and morbidity rate. In *Applied Change Point Problems in Statistics* (B. Sinha, A. Rukhin and M. Ahsanullah, eds.), 37–55.
- Mahdi Ebrahimzadeh, Noor A. Ibrahim, Abdul A. Jemain & Adem Kilicman. (2013). Claim dependence induced by common effects in hierarchical credibility models. *Communication in Statistics-Theory and Methods*, 42, 3373-3400 (DOI:10.1080/03610926.2011.625487).
- Mahdi Ebrahimzadeh, Noor A. Ibrahim, Abdul A. Jemain & Adem Kilicman (2012). Unbiased estimation of structural parameters in credibility models with dependence induced by common effects. *Bulletin of the Malaysian Mathematical Sciences Society* 35(2), 829-844.
- Mahdi Ebrahimzadeh, Noor Akma Ibrahim, Aziz Jemain, & Adem Kilicman. (2011). Credibility models with dependence induced by common effects and their applications in bank branch performance prediction. *Journal of International Business and Entrepreneurship Development*, 5(4), 339-350.
- Maller, R.A., & Zhou, X.(1996). *Survival analysis with long-term survivors*. New York: Wiley.
- McLachlan, G.J. & Krishnan, T. (2008). *The EM algorithm and extensions* (Second Edition). Wiley: Hoboken, New Jersey.
- Matthews, D. E., & Farewell, V. T. (1982). On testing for constant hazard against a change point alternative. *Biometrics*, 38, 463-468.
- Megan Othus, Yi Lib & Ram Tiwarid.(2012). Change-point cure models with application to estimating the change-point effect of age of diagnosis among prostate cancer patients. *Journal of Applied Statistics*, 39(4), 901–911.

- McCullagh, P., & J. A. Nelder. (1989).. *Generalized linear models* (2<sup>nd</sup> Edition). London: Chapman and Hall.
- Muller, H. G., & Wang, J. L., (1990). Nonparametric analysis of changes in hazard rates for censored survival data: an alternative to change-point models. *Biometrika*, 77, 305-314.
- N. A. Ibrahim, A. Kudus, I. Daud & M.R. Abu Bakar (2008). Decision tree for competing risks survival probability in breast cancer study. *International Journal of Biomedical Sciences*, 3(1), 25-29.
- Noor Akma Ibrahim, Abdul Kudus, Mohd. Rizam Abu Bakar & Isa Daud (2009). Cutpoint determination methods in competing risks subdistribution model. *Journal of Quality Management & Analysis JQMA*, 5(1), 103-117.
- Noor Akma Ibrahim & Abdul Kudus (2009). *Decision tree for prognostic classification of multivariate survival data and competing risks*. Chap1 :1-33. Recent Advances in Technologies, Edited by Maurizio A. Strangio, In-Teh, Croatia.
- Noor Akma Ibrahim & Suliadi (2011). Generating correlated discrete ordinal data using R and SAS IML. *Computer Methods and Program in Biomedicine*, 104(3), e122-e132.
- Noor Akma Ibrahim & Suliadi. (2009). Nonparametric regression for correlated data. *WSEAS Transactions on Mathematics*, 7(8), 353-362.
- Odell, P. M., Anderson, K. M., & D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a weibull-based accelerated failure time model. *Biometrics*. 48, 951-959.
- Pan, W.(2000). Multiple imputation approach to Cox regression with interval censored data. *Biometrics*, 56,199-203.
- Peng, Y., & Dear, K. (2000). A nonparametric mixture model for cure rate estimation, *Biometrics*, 56, 237-243.
- Pons, O. (2003). Estimation in a cox regression model with a change-point according to a threshold in a covariate. *Annals of Statistics*, 31, 442-463.
- Quinlan, J. R. (1992). C4.5: Program for machine learning. California: Morgan Kaufmann,

- Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.
- Segal, M. R. (1988). Regression trees for censored data. *Biometrics*, 44, 35-47.
- Segal, M. R. (1992). Tree-structured method for longitudinal data. *Journal of the American Statistical Association*, 87, 407-418.
- Segal, M.R. (1995). Extending the elements of tree-structured regression. *Statist. Methods Med. Res.*, 4, 219-236.
- Sen, P. K. (1993). Some change-point problems in survival analysis: relevance of nonparametric in applications. *Applied Change Point Problems in Statistics*, 325-336, Baltimore, MD.
- Suliadi Suliadi, Noor Akma Ibrahim & Isa Daud (2013). Semiparametric estimation with profile algorithm for longitudinal binary data. *Communications in Statistics – Simulation and Computation*, 42(1),138-152.
- Suliadi, Noor Akma Ibrahim, Isa Daud & Isthriyaygy S. Krishnarajah (2010a). Nonparametric Regression for Longitudinal Binary Data Based on GEE-Smoothing Spline. *Journal of Applied Probability and Statistics*, 5, (1), 77-93.
- Suliadi, Noor Akma Ibrahim, Isa Daud & Isthriyaygy S. Krishnarajah. (2010b). GEE-smoothing spline for semiparametric estimation of longitudinal binary data. *International Journal of Applied Mathematics and Statistics*, 18(S10), Sept 2010:82-95.
- Suliadi, Noor Akma Ibrahim, Isthriyaygy S. Krishnarajah and Isa Daud (2009). Comparison of Some Smoothing Parameter Selection Methods in Generalized Estimating Equation-Smoothing Spline. *World Applied Science Journal*, 7(11),1382-1387.
- Su, X. G., Fan, J. J. (2004). Multivariate survival trees: a maximum likelihood approach based on frailty models. *Biometrics*, 60, 93-99.
- Sun, J., Zhao, Q., & Zhao, X. (2005). Generalized log-rank tests for interval-censored failure time data. *Scand. J. Stat.*, 32, 49-57.

- Tsodikov, A.D., Ibrahim, J.G., & Yakovlev, A.Y.(2003). Estimating cure rates from survival data: An alternative to two-component mixture models, *J. Am. Stat. Assoc.*, 98, 1063–1078.
- Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 439-447.
- Welsh, Alan H., Xihong Lin, & Raymond J. Carroll. (2002). Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel methods. *Journal of the American Statistical Association*, 97, 482-493.
- Wu, Hulin & Jin-Ting Zhang.(2006). Nonparametric regression methods for longitudinal data analysis. John Wiley & Sons. USA.
- Xiaobing Zhao, Xianyi Wu & Xian Zhou. (2009). A change-point model for survival data with long-term survivors. *Statistica Sinica*, 19, 377-390.
- Yu, B., Tiwari, R.C., Cronin, K.A., & Feuer, E.J.(2004). Cure fraction estimation from the mixture cure models for grouped survival data. *Statistics in Medicine*, 23(11), 1733-1747.
- Zalina Mohd. Ali, Noor Akma Ibrahim, Kerrie mergesen, Mahendran Shitan, Hafizan Juahir & Faridatul Azna Ahmad Shahabuddin. (2012). *Temporal water quality assessment of Langat River from 1995-2006, Water Quality Monitoring and Assessment*, Intech, 321-346.
- Zeger, Scott L., & Kung-Yee Liang. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121-130.
- Zhang, H. P. (1998). Classification tree for multiple binary responses. *Journal of the American Statistical Association*, 93, 180-193.
- Zhao, X., Zhao Q., Sun, J., Kim, J. S. (2008), Generalized log-rank tests for partly interval- censored failure time data. *Biom. J.*, 50, 375–385.



## BIOGRAPHY

**Noor Akma Ibrahim** was born in Tapah, Perak. She received her early education in Johor Bahru and Kuala Kangsar and later on in Sungai Petani. In 1976 she received a scholarship award from Majlis Amanah Rakyat (MARA) and went on to pursue her first degree at Western Michigan University, USA. She was conferred the Bachelor of Science degree in Mathematics Finance in 1979 and Master of Science in Applied Statistics from the same university in 1982. In 1994 she obtained her Ph.D in Statistics from Universiti Putra Malaysia after joining the Department of Mathematics in 1982.

Noor Akma is currently a Professor in Statistics specializing in the area of Survival Analysis at the Department of Mathematics, Universiti Putra Malaysia. Her other research interest encompasses diagnostics and inference and computational methods in statistics. She headed the laboratory of Statistics and Applied Mathematics, Institute for Mathematical Research (INSPEM) from 2002-2004 and then elevated to the position of Deputy Director until 2012. She was a member of UPM Senate, appointed in 2011.

During her tenure years as a scholar, she has received several accolades including Institute for Mathematical Research Figure in 2010, Silver and Bronze medals for the year 2010 and 2006 respectively at the International Invention, Innovation, Industrial Design & Technology Exhibition (I-TEX), Malaysian Mathematical Sciences Society (PERSAMA) Main Award for the category of Mathematical Invention (2010) and a Highly Commended Award for the article published in Management Of Environmental Quality, an Emerald Literati Network journal (2006). Most of her research works were and are still funded by grants received from the Ministry of Science, Technology and Innovation (MOSTI), Ministry of Higher Education (MOHE) and UPM. Her collaboration with Dr. Abdul Kudus landed them a research grant from Third World

Academy of Science (TWAS). Nine PhD students and four Masters students had graduated under her supervision while seven are still active.

Her professional services include being an editorial board member of Malaysian Journal of Mathematical Sciences (MJMS), Journal of Quantitative Methods, International Journal for Applied Physics and Mathematics, KALAM and Journal of Data Analysis and Operational Research. She was also an editorial board member for PERTANIKA, Journal of Science & Technology and Chief Editor for InfoSains (1998-2002), a newsletter of the Faculty of Science. She has been invited to be a Reviewer by Math Review, American Mathematical Society. Her active affiliation in the Society include being a life member of the Malaysian Mathematical Sciences Society, Malaysian Academy of Mathematical Scientist (AISMM), elected as the Secretary, Management Science & Operational Research Society of Malaysia, Islamic Society of Statistical Sciences (ISOSS) and a member of Malaysian Institute of Statistics (ISM) and American Statistical Association (ASA). She has reviewed many papers for local and international journals including ISI journals and appointed as a panel for new program for both public and private universities. She has also served as a subject matter expert and examiner for the Malaysian Matriculation Centre.

Prof. Dr. Noor Akma was also involved in the Community program for the Orang Asli for two consecutive years spearheaded by INSPEM together with the Department of the Development of Orang Asli (JAKOA). Throughout her thirty-year stint at UPM, she has shown her mettle in her commitment in imparting knowledge and carrying out research for the good of mankind.

## **ACKNOWLEDGMENT**

I am most grateful to Allah SWT for His Blessings and Guidance. First and foremost I would like to thank Universiti Putra Malaysia, Ministry of Science, Technology and Innovation, and Ministry of Higher Learning for funding my research projects. My sincere gratitude goes to my supervisors, co-researchers, students and colleagues for their grateful thoughts and inspiration. I would like to extend this acknowledgment to the staff of Faculty of Science and Institute for Mathematical Research for their assistance and support. My utmost appreciation goes to my parents and my brothers and families for their prayers and love. Last but not least my undivided love and gratitude to my husband, Idris, my children, Liza and husband Hilal, Azlan, Adnan and Azlina for their endless unwavering love, support and sacrifice.

May Allah SWT grant all of us happiness in this world and in the hereafter.



## LIST OF INAUGURAL LECTURES

1. Prof. Dr. Sulaiman M. Yassin  
*The Challenge to Communication Research in Extension*  
22 July 1989
2. Prof. Ir. Abang Abdullah Abang Ali  
*Indigenous Materials and Technology for Low Cost Housing*  
30 August 1990
3. Prof. Dr. Abdul Rahman Abdul Razak  
*Plant Parasitic Nematodes, Lesser Known Pests of Agricultural Crops*  
30 January 1993
4. Prof. Dr. Mohamed Suleiman  
*Numerical Solution of Ordinary Differential Equations: A Historical Perspective*  
11 December 1993
5. Prof. Dr. Mohd. Ariff Hussein  
*Changing Roles of Agricultural Economics*  
5 March 1994
6. Prof. Dr. Mohd. Ismail Ahmad  
*Marketing Management: Prospects and Challenges for Agriculture*  
6 April 1994
7. Prof. Dr. Mohamed Mahyuddin Mohd. Dahan  
*The Changing Demand for Livestock Products*  
20 April 1994
8. Prof. Dr. Ruth Kiew  
*Plant Taxonomy, Biodiversity and Conservation*  
11 May 1994
9. Prof. Ir. Dr. Mohd. Zohadie Bardaie  
*Engineering Technological Developments Propelling Agriculture into the 21st Century*  
28 May 1994
10. Prof. Dr. Shamsuddin Jusop  
*Rock, Mineral and Soil*  
18 June 1994

### Natural Products from Malaysian Rainforests

11. Prof. Dr. Abdul Salam Abdullah  
*Natural Toxicants Affecting Animal Health and Production*  
29 June 1994
12. Prof. Dr. Mohd. Yusof Hussein  
*Pest Control: A Challenge in Applied Ecology*  
9 July 1994
13. Prof. Dr. Kapt. Mohd. Ibrahim Haji Mohamed  
*Managing Challenges in Fisheries Development through Science and Technology*  
23 July 1994
14. Prof. Dr. Hj. Amat Juhari Moain  
*Sejarah Keagungan Bahasa Melayu*  
6 Ogos 1994
15. Prof. Dr. Law Ah Theem  
*Oil Pollution in the Malaysian Seas*  
24 September 1994
16. Prof. Dr. Md. Nordin Hj. Lajis  
*Fine Chemicals from Biological Resources: The Wealth from Nature*  
21 January 1995
17. Prof. Dr. Sheikh Omar Abdul Rahman  
*Health, Disease and Death in Creatures Great and Small*  
25 February 1995
18. Prof. Dr. Mohamed Shariff Mohamed Din  
*Fish Health: An Odyssey through the Asia - Pacific Region*  
25 March 1995
19. Prof. Dr. Tengku Azmi Tengku Ibrahim  
*Chromosome Distribution and Production Performance of Water Buffaloes*  
6 May 1995
20. Prof. Dr. Abdul Hamid Mahmood  
*Bahasa Melayu sebagai Bahasa Ilmu- Cabaran dan Harapan*  
10 Jun 1995

Noor Akma Ibrahim

21. Prof. Dr. Rahim Md. Sail  
*Extension Education for Industrialising Malaysia: Trends, Priorities and Emerging Issues*  
22 July 1995
22. Prof. Dr. Nik Muhammad Nik Abd. Majid  
*The Diminishing Tropical Rain Forest: Causes, Symptoms and Cure*  
19 August 1995
23. Prof. Dr. Ang Kok Jee  
*The Evolution of an Environmentally Friendly Hatchery Technology for Udang Galah, the King of Freshwater Prawns and a Glimpse into the Future of Aquaculture in the 21st Century*  
14 October 1995
24. Prof. Dr. Sharifuddin Haji Abdul Hamid  
*Management of Highly Weathered Acid Soils for Sustainable Crop Production*  
28 October 1995
25. Prof. Dr. Yu Swee Yean  
*Fish Processing and Preservation: Recent Advances and Future Directions*  
9 December 1995
26. Prof. Dr. Rosli Mohamad  
*Pesticide Usage: Concern and Options*  
10 February 1996
27. Prof. Dr. Mohamed Ismail Abdul Karim  
*Microbial Fermentation and Utilization of Agricultural Bioresources and Wastes in Malaysia*  
2 March 1996
28. Prof. Dr. Wan Sulaiman Wan Harun  
*Soil Physics: From Glass Beads to Precision Agriculture*  
16 March 1996
29. Prof. Dr. Abdul Aziz Abdul Rahman  
*Sustained Growth and Sustainable Development: Is there a Trade-Off 1 or Malaysia*  
13 April 1996

Natural Products from Malaysian Rainforests

30. Prof. Dr. Chew Tek Ann  
*Sharecropping in Perfectly Competitive Markets: A Contradiction in Terms*  
27 April 1996
31. Prof. Dr. Mohd. Yusuf Sulaiman  
*Back to the Future with the Sun*  
18 May 1996
32. Prof. Dr. Abu Bakar Salleh  
*Enzyme Technology: The Basis for Biotechnological Development*  
8 June 1996
33. Prof. Dr. Kamel Ariffin Mohd. Atan  
*The Fascinating Numbers*  
29 June 1996
34. Prof. Dr. Ho Yin Wan  
*Fungi: Friends or Foes*  
27 July 1996
35. Prof. Dr. Tan Soon Guan  
*Genetic Diversity of Some Southeast Asian Animals: Of Buffaloes and Goats and Fishes Too*  
10 August 1996
36. Prof. Dr. Nazaruddin Mohd. Jali  
*Will Rural Sociology Remain Relevant in the 21st Century?*  
21 September 1996
37. Prof. Dr. Abdul Rani Bahaman  
*Leptospirosis-A Model for Epidemiology, Diagnosis and Control of Infectious Diseases*  
16 November 1996
38. Prof. Dr. Marziah Mahmood  
*Plant Biotechnology - Strategies for Commercialization*  
21 December 1996
39. Prof. Dr. Ishak Hj. Omar  
*Market Relationships in the Malaysian Fish Trade: Theory and Application*  
22 March 1997

Noor Akma Ibrahim

40. Prof. Dr. Suhaila Mohamad  
*Food and Its Healing Power*  
12 April 1997
41. Prof. Dr. Malay Raj Mukerjee  
*A Distributed Collaborative Environment for Distance Learning Applications*  
17 June 1998
42. Prof. Dr. Wong Kai Choo  
*Advancing the Fruit Industry in Malaysia: A Need to Shift Research Emphasis*  
15 May 1999
43. Prof. Dr. Aini Ideris  
*Avian Respiratory and Immunosuppressive Diseases- A Fatal Attraction*  
10 July 1999
44. Prof. Dr. Sariah Meon  
*Biological Control of Plant Pathogens: Harnessing the Richness of Microbial Diversity*  
14 August 1999
45. Prof. Dr. Azizah Hashim  
*The Endomycorrhiza: A Futile Investment?*  
23 Oktober 1999
46. Prof. Dr. Noraini Abdul Samad  
*Molecular Plant Virology: The Way Forward*  
2 February 2000
47. Prof. Dr. Muhamad Awang  
*Do We Have Enough Clean Air to Breathe?*  
7 April 2000
48. Prof. Dr. Lee Chnoong Kheng  
*Green Environment, Clean Power*  
24 June 2000
49. Prof. Dr. Mohd. Ghazali Mohayidin  
*Managing Change in the Agriculture Sector: The Need for Innovative Educational Initiatives*  
12 January 2002

Natural Products from Malaysian Rainforests

50. Prof. Dr. Fatimah Mohd. Arshad  
*Analisis Pemasaran Pertanian di Malaysia: Keperluan Agenda  
Pembaharuan*  
26 Januari 2002
51. Prof. Dr. Nik Mustapha R. Abdullah  
*Fisheries Co-Management: An Institutional Innovation Towards  
Sustainable Fisheries Industry*  
28 February 2002
52. Prof. Dr. Gulam Rusul Rahmat Ali  
*Food Safety: Perspectives and Challenges*  
23 March 2002
53. Prof. Dr. Zaharah A. Rahman  
*Nutrient Management Strategies for Sustainable Crop Production in Acid  
Soils: The Role of Research Using Isotopes*  
13 April 2002
54. Prof. Dr. Maisom Abdullah  
*Productivity Driven Growth: Problems & Possibilities*  
27 April 2002
55. Prof. Dr. Wan Omar Abdullah  
*Immunodiagnosis and Vaccination for Brugian Filariasis: Direct Rewards  
from Research Investments*  
6 June 2002
56. Prof. Dr. Syed Tajuddin Syed Hassan  
*Agro-ento Bioinformation: Towards the Edge of Reality*  
22 June 2002
57. Prof. Dr. Dahlan Ismail  
*Sustainability of Tropical Animal-Agricultural Production Systems:  
Integration of Dynamic Complex Systems*  
27 June 2002
58. Prof. Dr. Ahmad Zubaidi Baharumshah  
*The Economics of Exchange Rates in the East Asian Countries*  
26 October 2002
59. Prof. Dr. Shaik Md. Noor Alam S.M. Hussain  
*Contractual Justice in Asean: A Comparative View of Coercion*  
31 October 2002

Noor Akma Ibrahim

60. Prof. Dr. Wan Md. Zin Wan Yunus  
*Chemical Modification of Polymers: Current and Future Routes for Synthesizing New Polymeric Compounds*  
9 November 2002
61. Prof. Dr. Annuar Md. Nassir  
*Is the KLSE Efficient? Efficient Market Hypothesis vs Behavioural Finance*  
23 November 2002
62. Prof. Ir. Dr. Radin Umar Radin Sohadi  
*Road Safety Interventions in Malaysia: How Effective Are They?*  
21 February 2003
63. Prof. Dr. Shamsher Mohamad  
*The New Shares Market: Regulatory Intervention, Forecast Errors and Challenges*  
26 April 2003
64. Prof. Dr. Han Chun Kwong  
*Blueprint for Transformation or Business as Usual? A Structural Perspective of the Knowledge-Based Economy in Malaysia*  
31 May 2003
65. Prof. Dr. Mawardi Rahmani  
*Chemical Diversity of Malaysian Flora: Potential Source of Rich Therapeutic Chemicals*  
26 July 2003
66. Prof. Dr. Fatimah Md. Yusoff  
*An Ecological Approach: A Viable Option for Aquaculture Industry in Malaysia*  
9 August 2003
67. Prof. Dr. Mohamed Ali Rajion  
*The Essential Fatty Acids-Revisited*  
23 August 2003
68. Prof. Dr. Azhar Md. Zain  
*Psychotherapy for Rural Malays - Does it Work?*  
13 September 2003

## Natural Products from Malaysian Rainforests

69. Prof. Dr. Mohd. Zamri Saad  
*Respiratory Tract Infection: Establishment and Control*  
27 September 2003
70. Prof. Dr. Jinap Selamat  
*Cocoa-Wonders for Chocolate Lovers*  
14 February 2004
71. Prof. Dr. Abdul Halim Shaari  
*High Temperature Superconductivity: Puzzle & Promises*  
13 March 2004
72. Prof. Dr. Yaakob Che Man  
*Oils and Fats Analysis - Recent Advances and Future Prospects*  
27 March 2004
73. Prof. Dr. Kaida Khalid  
*Microwave Aquametry: A Growing Technology*  
24 April 2004
74. Prof. Dr. Hasanah Mohd. Ghazali  
*Tapping the Power of Enzymes- Greening the Food Industry*  
11 May 2004
75. Prof. Dr. Yusof Ibrahim  
*The Spider Mite Saga: Quest for Biorational Management Strategies*  
22 May 2004
76. Prof. Datin Dr. Sharifah Md. Nor  
*The Education of At-Risk Children: The Challenges Ahead*  
26 June 2004
77. Prof. Dr. Ir. Wan Ishak Wan Ismail  
*Agricultural Robot: A New Technology Development for Agro-Based Industry*  
14 August 2004
78. Prof. Dr. Ahmad Said Sajap  
*Insect Diseases: Resources for Biopesticide Development*  
28 August 2004

Noor Akma Ibrahim

79. Prof. Dr. Aminah Ahmad  
*The Interface of Work and Family Roles: A Quest for Balanced Lives*  
11 March 2005
80. Prof. Dr. Abdul Razak Alimon  
*Challenges in Feeding Livestock: From Wastes to Feed*  
23 April 2005
81. Prof. Dr. Haji Azimi Hj. Hamzah  
*Helping Malaysian Youth Move Forward: Unleashing the Prime Enablers*  
29 April 2005
82. Prof. Dr. Rasedee Abdullah  
*In Search of An Early Indicator of Kidney Disease*  
27 May 2005
83. Prof. Dr. Zulkifli Hj. Shamsuddin  
*Smart Partnership: Plant-Rhizobacteria Associations*  
17 June 2005
84. Prof. Dr. Mohd Khanif Yusop  
*From the Soil to the Table*  
1 July 2005
85. Prof. Dr. Annuar Kassim  
*Materials Science and Technology: Past, Present and the Future*  
8 July 2005
86. Prof. Dr. Othman Mohamed  
*Enhancing Career Development Counselling and the Beauty of Career Games*  
12 August 2005
87. Prof. Ir. Dr. Mohd Amin Mohd Soom  
*Engineering Agricultural Water Management Towards Precision Framing*  
26 August 2005
88. Prof. Dr. Mohd Arif Syed  
*Bioremediation-A Hope Yet for the Environment?*  
9 September 2005

Natural Products from Malaysian Rainforests

89. Prof. Dr. Abdul Hamid Abdul Rashid  
*The Wonder of Our Neuromotor System and the Technological Challenges They Pose*  
23 December 2005
90. Prof. Dr. Norhani Abdullah  
*Rumen Microbes and Some of Their Biotechnological Applications*  
27 January 2006
91. Prof. Dr. Abdul Aziz Saharee  
*Haemorrhagic Septicaemia in Cattle and Buffaloes: Are We Ready for Freedom?*  
24 February 2006
92. Prof. Dr. Kamariah Abu Bakar  
*Activating Teachers' Knowledge and Lifelong Journey in Their Professional Development*  
3 March 2006
93. Prof. Dr. Borhanuddin Mohd. Ali  
*Internet Unwired*  
24 March 2006
94. Prof. Dr. Sundararajan Thilagar  
*Development and Innovation in the Fracture Management of Animals*  
31 March 2006
95. Prof. Dr. Zainal Aznam Md. Jelani  
*Strategic Feeding for a Sustainable Ruminant Farming*  
19 May 2006
96. Prof. Dr. Mahiran Basri  
*Green Organic Chemistry: Enzyme at Work*  
14 July 2006
97. Prof. Dr. Malik Hj. Abu Hassan  
*Towards Large Scale Unconstrained Optimization*  
20 April 2007
98. Prof. Dr. Khalid Abdul Rahim  
*Trade and Sustainable Development: Lessons from Malaysia's Experience*  
22 Jun 2007

Noor Akma Ibrahim

99. Prof. Dr. Mad Nasir Shamsudin  
*Econometric Modelling for Agricultural Policy Analysis and Forecasting: Between Theory and Reality*  
13 July 2007
100. Prof. Dr. Zainal Abidin Mohamed  
*Managing Change - The Fads and The Realities: A Look at Process Reengineering, Knowledge Management and Blue Ocean Strategy*  
9 November 2007
101. Prof. Ir. Dr. Mohamed Daud  
*Expert Systems for Environmental Impacts and Ecotourism Assessments*  
23 November 2007
102. Prof. Dr. Saleha Abdul Aziz  
*Pathogens and Residues; How Safe is Our Meat?*  
30 November 2007
103. Prof. Dr. Jayum A. Jawan  
*Hubungan Sesama Manusia*  
7 Disember 2007
104. Prof. Dr. Zakariah Abdul Rashid  
*Planning for Equal Income Distribution in Malaysia: A General Equilibrium Approach*  
28 December 2007
105. Prof. Datin Paduka Dr. Khatijah Yusoff  
*Newcastle Disease virus: A Journey from Poultry to Cancer*  
11 January 2008
106. Prof. Dr. Dzulkefly Kuang Abdullah  
*Palm Oil: Still the Best Choice*  
1 February 2008
107. Prof. Dr. Elias Saion  
*Probing the Microscopic Worlds by Ionizing Radiation*  
22 February 2008
108. Prof. Dr. Mohd Ali Hassan  
*Waste-to-Wealth Through Biotechnology: For Profit, People and Planet*  
28 March 2008

Natural Products from Malaysian Rainforests

109. Prof. Dr. Mohd Maarof H. A. Moxsin  
*Metrology at Nanoscale: Thermal Wave Probe Made It Simple*  
11 April 2008
110. Prof. Dr. Dzolkhifli Omar  
*The Future of Pesticides Technology in Agriculture: Maximum Target Kill with Minimum Collateral Damage*  
25 April 2008
111. Prof. Dr. Mohd. Yazid Abd. Manap  
*Probiotics: Your Friendly Gut Bacteria*  
9 May 2008
112. Prof. Dr. Hamami Sahri  
*Sustainable Supply of Wood and Fibre: Does Malaysia have Enough?*  
23 May 2008
113. Prof. Dato' Dr. Makhdzir Mardan  
*Connecting the Bee Dots*  
20 June 2008
114. Prof. Dr. Maimunah Ismail  
*Gender & Career: Realities and Challenges*  
25 July 2008
115. Prof. Dr. Nor Aripin Shamaan  
*Biochemistry of Xenobiotics: Towards a Healthy Lifestyle and Safe Environment*  
1 August 2008
116. Prof. Dr. Mohd Yunus Abdullah  
*Penjagaan Kesihatan Primer di Malaysia: Cabaran Prospek dan Implikasi dalam Latihan dan Penyelidikan Perubatan serta Sains Kesihatan di Universiti Putra Malaysia*  
8 Ogos 2008
117. Prof. Dr. Musa Abu Hassan  
*Memfaatkan Teknologi Maklumat & Komunikasi ICT untuk Semua*  
15 Ogos 2008
118. Prof. Dr. Md. Salleh Hj. Hassan  
*Role of Media in Development: Strategies, Issues & Challenges*  
22 August 2008

Noor Akma Ibrahim

119. Prof. Dr. Jariah Masud  
*Gender in Everyday Life*  
10 October 2008
120. Prof. Dr. Mohd Shahwahid Haji Othman  
*Mainstreaming Environment: Incorporating Economic Valuation and Market-Based Instruments in Decision Making*  
24 October 2008
121. Prof. Dr. Son Radu  
*Big Questions Small Worlds: Following Diverse Vistas*  
31 Oktober 2008
122. Prof. Dr. Russly Abdul Rahman  
*Responding to Changing Lifestyles: Engineering the Convenience Foods*  
28 November 2008
123. Prof. Dr. Mustafa Kamal Mohd Shariff  
*Aesthetics in the Environment an Exploration of Environmental: Perception Through Landscape Preference*  
9 January 2009
124. Prof. Dr. Abu Daud Silong  
*Leadership Theories, Research & Practices: Farming Future Leadership Thinking*  
16 January 2009
125. Prof. Dr. Azni Idris  
*Waste Management, What is the Choice: Land Disposal or Biofuel?*  
23 January 2009
126. Prof. Dr. Jamilah Bakar  
*Freshwater Fish: The Overlooked Alternative*  
30 January 2009
127. Prof. Dr. Mohd. Zobir Hussein  
*The Chemistry of Nanomaterial and Nanobiomaterial*  
6 February 2009
128. Prof. Ir. Dr. Lee Teang Shui  
*Engineering Agricultural: Water Resources*  
20 February 2009

Natural Products from Malaysian Rainforests

129. Prof. Dr. Ghizan Saleh  
*Crop Breeding: Exploiting Genes for Food and Feed*  
6 March 2009
130. Prof. Dr. Muzafar Shah Habibullah  
*Money Demand*  
27 March 2009
131. Prof. Dr. Karen Anne Crouse  
*In Search of Small Active Molecules*  
3 April 2009
132. Prof. Dr. Turiman Suandi  
*Volunteerism: Expanding the Frontiers of Youth Development*  
17 April 2009
133. Prof. Dr. Arbakariya Ariff  
*Industrializing Biotechnology: Roles of Fermentation and Bioprocess Technology*  
8 Mei 2009
134. Prof. Ir. Dr. Desa Ahmad  
*Mechanics of Tillage Implements*  
12 Jun 2009
135. Prof. Dr. W. Mahmood Mat Yunus  
*Photothermal and Photoacoustic: From Basic Research to Industrial Applications*  
10 Julai 2009
136. Prof. Dr. Taufiq Yap Yun Hin  
*Catalysis for a Sustainable World*  
7 August 2009
137. Prof. Dr. Raja Noor Zaliha Raja Abd. Rahman  
*Microbial Enzymes: From Earth to Space*  
9 Oktober 2009
138. Prof. Ir. Dr. Barkawi Sahari  
*Materials, Energy and CNGDI Vehicle Engineering*  
6 November 2009

Noor Akma Ibrahim

139. Prof. Dr. Zulkifli Idrus  
*Poultry Welfare in Modern Agriculture: Opportunity or Threat?*  
13 November 2009
140. Prof. Dr. Mohamed Hanafi Musa  
*Managing Phosphorus: Under Acid Soils Environment*  
8 January 2010
141. Prof. Dr. Abdul Manan Mat Jais  
*Haruan Channa striatus a Drug Discovery in an Agro-Industry Setting*  
12 March 2010
142. Prof. Dr. Bujang bin Kim Huat  
*Problematic Soils: In Search for Solution*  
19 March 2010
143. Prof. Dr. Samsinar Md Sidin  
*Family Purchase Decision Making: Current Issues & Future Challenges*  
16 April 2010
144. Prof. Dr. Mohd Adzir Mahdi  
*Lightspeed: Catch Me If You Can*  
4 June 2010
145. Prof. Dr. Raha Hj. Abdul Rahim  
*Designer Genes: Fashioning Mission Purposed Microbes*  
18 June 2010
146. Prof. Dr. Hj. Hamidon Hj. Basri  
*A Stroke of Hope, A New Beginning*  
2 July 2010
147. Prof. Dr. Hj. Kamaruzaman Jusoff  
*Going Hyperspectral: The "Unseen" Captured?*  
16 July 2010
148. Prof. Dr. Mohd Sapuan Salit  
*Concurrent Engineering for Composites*  
30 July 2010
149. Prof. Dr. Shattri Mansor  
*Google the Earth: What's Next?*  
15 October 2010

Natural Products from Malaysian Rainforests

150. Prof. Dr. Mohd Basyaruddin Abdul Rahman  
*Haute Couture: Molecules & Biocatalysts*  
29 October 2010
151. Prof. Dr. Mohd. Hair Bejo  
*Poultry Vaccines: An Innovation for Food Safety and Security*  
12 November 2010
152. Prof. Dr. Umi Kalsom Yusuf  
*Fern of Malaysian Rain Forest*  
3 December 2010
153. Prof. Dr. Ab. Rahim Bakar  
*Preparing Malaysian Youths for The World of Work: Roles of Technical  
and Vocational Education and Training (TVET)*  
14 January 2011
154. Prof. Dr. Seow Heng Fong  
*Are there "Magic Bullets" for Cancer Therapy?*  
11 February 2011
155. Prof. Dr. Mohd Azmi Mohd Lila  
*Biopharmaceuticals: Protection, Cure and the Real Winner*  
18 February 2011
156. Prof. Dr. Siti Shapor Siraj  
*Genetic Manipulation in Farmed Fish: Enhancing Aquaculture Production*  
25 March 2011
157. Prof. Dr. Ahmad Ismail  
*Coastal Biodiversity and Pollution: A Continuous Conflict*  
22 April 2011
158. Prof. Ir. Dr. Norman Mariun  
*Energy Crisis 2050? Global Scenario and Way Forward for Malaysia*  
10 June 2011
159. Prof. Dr. Mohd Razi Ismail  
*Managing Plant Under Stress: A Challenge for Food Security*  
15 July 2011

Noor Akma Ibrahim

160. Prof. Dr. Patimah Ismail  
*Does Genetic Polymorphisms Affect Health?*  
23 September 2011
161. Prof. Dr. Sidek Ab. Aziz  
*Wonders of Glass: Synthesis, Elasticity and Application*  
7 October 2011
162. Prof. Dr. Azizah Osman  
*Fruits: Nutritious, Colourful, Yet Fragile Gifts of Nature*  
14 October 2011
163. Prof. Dr. Mohd. Fauzi Ramlan  
*Climate Change: Crop Performance and Potential*  
11 November 2011
164. Prof. Dr. Adem Kiliçman  
*Mathematical Modeling with Generalized Function*  
25 November 2011
165. Prof. Dr. Fauziah Othman  
*My Small World: In Biomedical Research*  
23 December 2011
166. Prof. Dr. Japar Sidik Bujang  
*The Marine Angiosperms, Seagrass*  
23 March 2012
166. Prof. Dr. Zailina Hashim  
*Air Quality and Children's Environmental Health: Is Our Future  
Generation at Risk?*  
30 March 2012
167. Prof. Dr. Zainal Abidin Mohamed  
*Where is the Beef? Vantage Point form the Livestock Supply Chain*  
27 April 2012
168. Prof. Dr. Jothi Malar Panandam  
*Genetic Characterisation of Animal Genetic Resources for Sustainable  
Utilisation and Development*  
30 November 2012

Natural Products from Malaysian Rainforests

169. Prof. Dr. Fatimah Abu Bakar  
*The Good The Bad & Ugly of Food Safety: From Molecules to Microbes*  
7 Desember 2012
170. Prof. Dr. Abdul Jalil Nordin  
*My Colourful Sketches from Scratch: Molecular Imaging*  
5 April 2013
171. Prof. Dr. Norlijah Othman  
*Lower Respiratory Infections in Children: New Pathogens, Old Pathogens  
and The Way Forward*  
19 April 2013
172. Prof. Dr. Jayakaran Mukundan  
*Steroid-like Prescriptions English Language Teaching Can Ill-afford*  
26 April 2013
173. Prof. Dr. Azmi Zakaria  
*Photothermals Affect Our Lives*  
7 Jun 2013
174. Prof. Dr. Rahinah Ibrahim  
*Design Informatics*  
21 Jun 2103
175. Prof. Dr. Gwendoline Ee Cheng Lian  
*Natural Products from Malaysian Rainforests*  
1 November 2013