Statistical Approaches Applied on a Greek Lip Cancer Dataset

K. Kitikidou[*1], A. Ntomouchtsis[2], P. Xirou[2], K. Paraskevopoulos[2], K. Vahtsevanos[2], C. Tsompanidou[2], G. Koloutsos[2], B. Christoforidou[2], K. Kontos[2], G. Kechagias[2], K. Andoniades[2]

[1]Democritus University of Thrace, Greece

[2]Cancer Hospital Theagenion, Thessaloniki, Greece

[*1]kkitikid@fmenr.duth.gr

***Abstract*- Having at our disposal a dataset of 186 lip cancer cases in Greece, we attempt to interpret them by applying four different statistical methods: Generalized Linear Model (GLM), Markov Chain Monte Carlo (MCMC), Cox proportional hazards model and Bayes factor. The likelihood score equations from GLM exerted estimators with bounded influence, so that the resulting estimators were robust against outliers while maintaining high efficiency in the absence of outliers. Batch means method of estimating the variance of the asymptotic normal distribution, used in MCMC, gave strong consistency when it was applied to our data. A Cox proportional hazards model done with a weighted expectation-maximization gave efficient parameter estimates. Finally, using Bayes factor to the prior distributions for the parameters in compared regression models was proved to be highly sensitive.**

Keywords*- Bayes Factor; Cox Regression; Generalized Linear Model; Lip cancer; Markov Chain Monte Carlo

I. INTRODUCTION

One of the questions many people ask when first diagnosed with cancer is about their prognosis. Patients might want to know whether their cancer is relatively easy or more difficult to treat. A doctor cannot predict the future, but an estimate is possible based on medical data. The study of cancer patients' data can give an idea of prognosis — the chance of survival. Statistics can also show how people with the same cancer type and stage respond to treatment. Doctors can use this information to weigh the pros and cons of each treatment option and develop a treatment plan. Thus, the selection of an appropriate statistical method for survivability prediction is of great importance. General Linear Models, Markov Chain Monte Carlo methods, Cox regression and Bayesian statistics are four techniques, among others, used to find good ways to predict survivability of cancer patients.

A. Generalized Linear Model (GLM)

Generalized Linear Model (GLM) is very attractive for handling a wide variety of continuous and discrete dependent variables. The response variables y_i, for i=1, 2, …, n approximate an exponential distribution such as $E(y_i)= \mu_i$, $\text{var}(y_i)=V(\mu_i)$ and linear predictor,

$$n_i = g(\mu_i) = \mathbf{x}_i^t\boldsymbol{\beta}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of parameters and g is a link function. The Maximum Likelihood Estimators (MLE) for $\boldsymbol{\beta}$ can be obtained by solving the score equations:

$$\sum_{i=1}^{n} \frac{y_i - \mu_i \partial \mu_i}{V(\mu_i)\partial n_i}\mathbf{x}_i = 0 \quad (1)$$

These MLE have unbounded influence, so they are sensitive to outliers. Reference [1] developed a robust estimation for GLM, considering a general class of M-estimators of Mallows's type, which are the solution of the amended score equations,

$$\sum_{i=1}^{n}\left\{\psi_c(r_i)\frac{1}{V^{1/2}(\mu_i)}\frac{\partial \mu_i}{\partial n_i}\mathbf{x}_i - \mathbf{a}(\boldsymbol{\beta})\right\} = 0$$

Where $r_i=(y_i-\mu_i)/V1/2(\mu_i)$, $\psi_c(r_i)=\max(-c, \min(r_i,c))$ is the Huber function, giving an estimator with bounded influence and $\boldsymbol{\alpha}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} E(\psi_c(r_i))\frac{1}{V^{1/2}(\mu_i)}\frac{\partial \mu_i}{\partial n_i}\mathbf{x}_i$.

This approach has been studied by [2], [3], [4] and [5]. Extensions to Generalized Linear Mixed Model (GLMM) have been developed by [6] and [7]. Finally, [8] developed Poisson and binomial GLM and their extensions to GLMM.

DOI：10.5963/LSMR0304003

B. Markov Chain Monte Carlo (MCMC)

Suppose that we want to calculate $E_\pi g = \int \chi g(x)\pi(dx)$, with π a probability distribution having support χ and g a real-valuated, π-integrable function. Suppose that π is such that Markov chain Monte Carlo (MCMC) is the only viable method for estimating $E_\pi g$.

Let $X=\{X_0, X_1, \ldots\}$ be a time-homogeneous, aperiodic, π-irreducible, positive Harris recurrent Markov chain [9]. In this case, X is the Harris ergodic and given an MCMC algorithm that simulates X, it is conceptually easy to generate large amounts of data and use $\bar{g}_n$ to obtain an arbitrarily precise estimate of $E_\pi g$.

Several methods can be used to determine when n is sufficiently large, i.e. when to terminate the simulation ([10]; [11]; [12]; [13]). We consider two methods for estimating the variance of the asymptotic normal distribution: Regenerative Simulation (RS) and nonoverlapping Batch Means (BM). Both have strengths and weaknesses: BM is easier to implement, while RS is on a stronger theoretical footing. Conditions provided for the consistency of BM, allowing the batch sizes to increase as n increases, are given by [14]. In this case, we denoted the method as CBM to distinguish it from the standard fixed-batch size version.

C. Cox Regression

In clinical trials and observatory studies, complete covariate data are often not available for every subject. Missing data may arise due to many circumstances, including the loss of hospital records or survey nonresponse. Intuitively, when the subjects with missing covariates differ systematically from those with complete data with respect to the outcome of interest, results from a traditional data analysis that omits the missing cases may no longer be valid.

Because standard techniques of survival analysis require full covariate information, a simple way to avoid the problem of missing data is to analyze only those subjects who are completely observed (complete case analysis). However, complete case analysis can be biased when the data are not Missing At Random (MAR) and generally leads to large standard errors. In addition, as the fraction of missing data increases, the deletion of all subjects with missing data is wasteful and inefficient.

Another method for dealing with missing covariates is to exclude those covariates subject to missingness from the analysis. However, this procedure leads to model misspecification. A method proposed by [15] uses an Expectation-Maximization (EM) Monte Carlo algorithm to obtain parameter estimates in the Cox proportional hazards model in the presence of missing categorical covariates. Reference [16] propose a different approximation that uses a weighted EM algorithm to obtain parameter estimates when the missing covariates are categorical or continuous. Previous work in this area includes methods developed by [17], [18], [19], [20], [21], [22], [23], [24].

D. Bayes factor

The Bayes factor is a tool for model selection for Bayesian statisticians. The Bayes factor summarizes the evidence provided by the data in favor of one scientific theory represented by a statistical model as opposed to another. Reference [25] provided a comprehensive review of Bayes factor including information about their interpretation. Bayes factor is known to be highly sensitive to the prior distributions used on the parameters of the models. Hence it is important to study the sensitivity of the Bayes factor to the prior distributions before drawing any conclusions. The traditional approach to study the sensitivity of the Bayes factor to the prior distribution is to evaluate the Bayes factor over classes of prior distributions [25]. However, this becomes very intensive computationally and there is no unanimity regarding the question of what class should be used. Reference [26] proposed an approach for studying the sensitivity of Bayes factor without the need of using a specific class of prior distributions.

The aims of this article are to:

1. investigate the robustness of GLM and make comparisons with other robust estimators,

2. examine the finite-sample properties of batch means methods (MCMC),

3. use a weighted expectation-maximization in a Cox proportional hazards model to handle missing covariate data, and

4. use Bayes factor, a Bayesian statistician's tool, for model selection, by applying these statistical methods to a dataset of lip cancer rates, in order to improve survivability prediction of lip cancer patients.

II. MATERIALS AND METHODS

Reference [27] analyzed n=186 lip cancer cases in Greece.

DOI：10.5963/LSMR0304003

A. Generalized Linear Model (GLM)

Consider a hierarchical model $y_i|\sigma_i \sim \mathrm{N}(\mu_t, \sigma_i^2)$, where $\mu_i = E(y_i|\sigma_i) = \mathbf{x}_i^t\boldsymbol{\beta}$. This leads to the model $y_i = \mathbf{x}_i^t\boldsymbol{\beta} + e_i$ (1) where $e_i=\sigma_i z_i$ and $z_i \sim \mathrm{N}(0,1)$. When $\sigma_i^2 = \sigma^2$ we have the ordinary regression model.

Suppose that conditional on unobservable random variables φ_i for i=1, 2, …, n, y_i follows a negative-binomial distribution with density function

$$f(y_i|\varphi_t) = \frac{\Gamma(y_i + 1/\varphi_t)}{y_i!\Gamma(1/\varphi_t)}\left(\frac{\varphi_i\mu_i}{1+\varphi_i\mu_i}\right)^{y_i}(1+\varphi_i\mu_i)^{-1/\varphi_i}.$$

Let $\varphi_i=\tau u_i$, where u_i follows the inverse-gamma distribution.

Here $\mathrm{var}(y_i|\varphi_i) = \mu_i(1+\varphi_i\mu_i)$. When τ=0, it becomes the Poisson model, with estimating equation (1) for $\boldsymbol{\beta}$ with $V(\mu)=\mu$; this gives unbounded influences as y_i goes to infinity.

Suppose that, conditional on φ_i, y_i follows a beta-binomial distribution with density function

$$f(y_i|\varphi_t) = \binom{m_i}{y_i}\prod_{k=0}^{y_i-1}(p_i+\varphi_i)\times\prod_{k=0}^{m_i-y_i-1}(1-p_i+k\varphi_i)\Big/\prod_{k=0}^{m_i-1}(1+k\varphi_i)$$

Here $\mathrm{var}(y_i|\varphi_i) = m_i p_i(1-p_i)\times\{1+(m_i-1)\varphi_i/(1+\varphi_i)\}$. When τ=0, it becomes the binomial model with estimating equation (1) for $\boldsymbol{\beta}$ with $V(\mu)=mp(1\text{-}p)$; this gives unbounded influences when y_i approaches 0 (or m_i), with m_i going to infinity.

In the Poisson GLMM, the h likelihood estimator $\tilde{u}_i$ for u_i satisfies the relation $\tilde{u}_i = O(y_{i+})$, with $y_{i+} = \sum_j y_{ij}$, as y_i+ goes to infinity and in the binomial GLMM, the h likelihood estimator $\tilde{u}_i$ for u_i satisfies the relation $\tilde{u}_i = O(m_{i+})$, with $m_{i+} = \sum_j m_{ij}$, as y_i+ approaches to 0 or m_i+, with m_i+ going to infinity (Noh and Lee 2007).

In our study, there was an effort to avoid the presentation of unstable rates for the smaller ages. For analysis, the following model was considered [28]:

$$n_i = \log\mu_i = \log n_i + \beta_0 + \beta_1 x_i/10 + u_i$$

where $\mathbf{v} = (u_1, u_2, ...u_n)^t \sim \exp\left\{-\sum_{i\sim j}(u_i-u_j)2/2\lambda\right\}$ and $i \sim j$ denotes adjacent ages.

B. Markov Chain Monte Carlo (MCMC)

Consider the Greek lip cancer dataset consisting of 186 lip cancer cases registered [27], together with the expected number of cases given the age-sex structure of the population. We assume a Poisson likelihood for spatially aggregated data. Specifically, for i=1, 2, …, N, we assume that given μ_i, the disease counts y_i are conditionally independent and $y_i|\mu_i \sim Poisson(E_i e^{\mu_i})$ where E_i is the expected number of disease events, assuming constant risk and μ_i is the log-relative risk of disease for the ith region. Set $\varphi=(\varphi_1, \ldots, \varphi_N)^T$. Each μ_i is modeled as $\mu_i=\theta_i+\varphi_i$, where $\theta_i|\tau_h \sim N(0,1/\tau_h)$, $\theta_i|\tau_c \sim CAR(\tau_c) \propto \tau_c^{N/2}\exp\left(-\frac{\tau_c}{2}\varphi^T Q\varphi\right)$ and $Q_{ij} = \begin{cases} n_i \text{ if } i=j \\ 0 \text{ if } i \text{ is not adjacent to } j \\ -1 \text{ if } i \text{ is adjacent to } j \end{cases}$

where n_i is the number of neighbors for the ith region. Each θ_i captures the ith region's extra-Poisson variability due to area wide heterogeneity, whereas each φ_i captures the ith region's excess variability attributable to regional clustering. The priors on the precision parameters are $\tau_H \sim gamma(1, 0.02)$. It is important that the random-effects parameters (θ_i, φ_i) are not identified in the likelihood and the spatial prior used is improper.

Reference [29] established uniform ergodicity of a Harris-ergodic Metropolis-Hastings independence sampler with invariant distribution $\pi(\theta, \varphi, \tau_h, \tau_c|y)$ where $\theta=(\theta_1, \ldots, \theta_N)^T$. In our implementation of RS, we used the formula for the probability of a regeneration given by [30].

C. Cox Regression

Think on the Greek lip cancer dataset of 186 lip cancer cases[27]. We compare our method to estimation based on complete cases, because other methods valid under MAR do not handle mixed covariates with a nonmonotonic pattern of missingness.

This study involved n=186 patients with outcome of interest in the overall survival, defined as the time from randomization from receiving radiotherapy or observation (z_1) until death from squamous cell carcinoma. In addition, several prognostic factors were identified as important predictors of survival. These included the patient's age (z_2), thickness of the tumor in mm (z_3), size of the primary tumor in cm^2 (z_4) and the type of the primary tumor (z_5), which was either superficial spreading or

DOI：10.5963/LSMR0304003

other. Treatment and age were observed for all patients, while thickness, size and type were missing for some subjects, so that 30% of the patients had missing covariate data. Logarithms of age, thickness and size were used in all analyses to achieve approximate normality in the distributions of the continuous covariates, which were also standardized to have mean 0 and variance 1.

We assume that the missingness does not depend on the values of the missing covariates. We use Cox regression to model the relationship between overall survival and the given prognostic factors. Because three covariates are subject to missingness, we must specify their distribution for the analysis. Using the [15] model

$$p(z_{i1},...,z_{ir}|\boldsymbol{\alpha}) = p(z_{i1}|z_{i1},...,z_{i,r-1},\mathbf{v}_i,\boldsymbol{\alpha}_r)\times p(z_{i,r-1}|z_{i1},...,z_{i,r-2},\mathbf{v}_i,\boldsymbol{\alpha}_{r-1})...\times p(z_{i1}|\mathbf{v}_i,\boldsymbol{\alpha}_1)$$

where $\boldsymbol{\alpha}_j$ is a vector of location and scale parameters for the jth conditional distribution, the $\boldsymbol{\alpha}_j$'s are distinct and $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_r)$, we model the covariate distribution as

$$p(z_{i3},z_{i4},z_{i5}|z_{i1},z_{i2},\boldsymbol{\alpha}) = p(z_{i5}|z_{i1},z_{i2},z_{i3},z_{i4},\boldsymbol{\alpha}_5)\times p(z_{i4}|z_{i1},z_{i2},z_{i3},\boldsymbol{\alpha}_4)...\times p(z_{i3}|z_{i1},z_{i2},\boldsymbol{\alpha}_3)$$

where i=1, 2, …, n [16]. Because treatment and age are always observed, they do not need to be modeled and are conditioned upon through the analysis. We model tumor type, a dichotomous covariate, using a logistic regression model. We then model the continuous covariates size and thickness as normal random variables.

D. Bayes factor

Once again, consider the Greek lip cancer dataset [27]. Patterns of regional variation in the disease incidence rate for lip cancer were investigated. The dataset contains: the observed number of lip cancer cases among males from 1995-2005, y_1, y_2, …y_n; the expected number of cases adjusted for the age distribution E_1, E_2, …, E_n; the percentage of people employed in agriculture, fishing and construction (AFC) x_1, x_2, …, x_n (since increased exposure to sunlight has been implicated in the excess occurrence of lip cancer, people working outdoors were thought to be under greater risk of the disease); and the set of neighboring regions N_1, N_2, …, N_n.

Assume that the disease incidence counts y_i's follow the independent Poisson distributions

$$y_i|\lambda_i) \sim Poisson(\lambda_i, E_i)\ ,\ i=1, 2, \ldots, n,$$

with λ_i representing a relative risk parameter for the ith region. In practice the Poisson model by itself does not provide an adequate fit due to factors not included in the model. One way to address the effects of unobserved covariates is to incorporate random effects. Reference [31] propose a mixed linear model for the vector of log relative risk parameters, $\log(\lambda)$:

$$\log(\lambda) = \mathbf{X}\beta + n + \psi\ ,$$

where $\mathbf{X}$ is the covariate matrix containing a vector of 1's as the first column and another column containing values of the variable AFC; β is a vector containing the fixed effect parameters β_0 and β_1; $n=(n_1, n_2, \ldots, n_n)$' is a vector of spatially random effects and $\psi=(\psi_1, \psi_2, \ldots, \psi_n)$ is a vector of uncorrelated heterogeneity random effects.

The spatial random effects n_i's are intended to represent unobserved factors that, if observed, would display substantial spatial correlation in that the values for a pair of contiguous zones would be generally much more alike than for two arbitrary zones. For known matrices C and M, we take the prior distribution for n as a conditional autoregressive (CAR) distribution

$$n|\tau^2,\varphi \sim N(0,\tau^2(1-\varphi\mathbf{C})^{-1}\mathbf{M})$$

where τ^2 and φ are parameters of the prior distribution and $\mathbf{C}$, $\mathbf{M}$ matrices as suggested by [32].

We focus on the Bayes factor (BF) for comparing the Poisson regression model to the CAR model, in order to test if there is any spatial structure evident in the unexplained risk. The Bayes factor of interest BF^{10}_{φ,τ^2} is defined in [26].

III. RESULTS AND DISCUSSION

A. Generalized Linear Model (GLM)

The results of fitting Poisson and binomial GLMM are given in Table 1. Our robust results are similar to those from the MLE method without the outliers.

TABLE I SUMMARIES OF ANALYSIS FOR THE GREEK LIP CANCER DATA

	Classical MLE method				Poisson GLMM		Binomial GLMM	
	Full data		Without outliers					
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
β_0	-0.246	0.148	-0.394	0.160	-0.369	0.160	-0.381	0.160
β_1	0.431	0.160	0.566	0.160	0.541	0.148	0.541	0.160
λ	2.962		2.838		4.442		12.957	

DOI：10.5963/LSMR0304003

Considering the case study, the use of robust statistical models for GLM classes gives robust estimations against outliers while maintaining high efficiency in the absence of outliers.

B. Markov Chain Monte Carlo (MCMC)

Table 2 reveals that the estimates of the coverage probabilities are all below the desired 0.95, but BM has the lowest standard error. Coverages show that the chains do not mix as well. This is due to the nature of the data and hence larger batch sizes are required.

TABLE II SUMMARIES FOR BM, CBM AND RS

Method	b_n	*n*/R**	Average half-width	Average chain length	Coverage probability
CBM	$n^{1/2}$	10,000	0.004	160	0.93
CBM	$n^{1/3}$	10,000	0.004	130	0.90
BM	n/30	10,000	0.003	130	0.80
RS		25	0.004	170	0.94

Whereas CBM and RS appear comparable in terms of coverage probability, RS tends to result in slightly longer runs than CBM, which in turn results in longer runs than BM. Moreover, RS and CBM are comparable in their ability to produce intervals that meet the target half-width more closely than BM. Also, the intervals for RS are apparently more stable than those of CBM and BM. Finally, BM underestimates the Monte Carlo standard error and thus suggests stopping the chain too early.

In conclusion, RS and CBM appear comparable. Also, like RS, CBM avoids the burn-in issue, which has been an obstacle to MCMC practitioners. CBM is slightly easier to implement, thus CBM has a place in the tool kit of MCMC practitioners.

C. Cox Regression

The results of this analysis are compared to the traditional complete case analysis in Table 3. Both methods yield similar qualitative results in terms of the treatment effect. Patients who received radiotherapy survived longer than those who were only observed after surgery. However, in the complete case analysis, the treatment effect achieves only marginal statistical significance (p=0.06), whereas the effect of radiotherapy is statistically significant (p=0.02) in the proposed EM method once the additional information provided by partially observed patients is taken into account. In addition, the effect of age becomes marginally significant (p=0.10) in the EM analysis, indicating that older patients have a poorer prognosis. Thus for these data, we see that a complete case analysis indicates that treatment has no significant effect on survival at the 0.05 level, whereas the proposed EM analysis shows that treatment does significantly improve survival.

TABLE III ESTIMATES FOR SQUAMOUS CELL CARCINOMA DATA

Effect	Method	Estimate	Standard error	*p*-value
Treatment	Complete case	-0.34	0.17	0.06
	EM	-0.30	0.12	0.02
Age	Complete case	0.08	0.07	0.40
	EM	0.10	0.06	0.10
Tumor thickness	Complete case	-0.09	0.10	0.40
	EM	-0.01	0.07	0.90
Tumor size	Complete case	0.03	0.13	0.90
	EM	0.03	0.08	0.70
Tumor type	Complete case	-0.01	0.20	0.99
	EM	0,01	0,14	0,90

D. Bayes factor

The approximate Bayes factor obtained using

$$p(\varphi,\tau^2)=\frac{1}{\varphi_{\max}}I_{\varphi\in(0,\varphi_{\max})}\cdot Inv-gamma(\tau^2\,|\,2.5,1.5)$$

is 0.01, which seems to give a strong evidence against the null hypothesis indicating that there seems to be evidence of spatial correlation in the unexplained risk of lip cancer incidence in Greece. However, before making any definite conclusion, we should examine if the Bayes factor of interest is sensitive to the prior distribution $p(\varphi,\tau^2)$. The values of the quantity BF^{10}_{φ,τ^2} ("point mass prior Bayes factor"), comparing the Poisson regression model with the CAR model, are computed for a grid of values of φ and τ^2. The quantity of BF^{10}_{φ,τ^2} varied a great deal with change in φ and τ^2; in fact, on logarithmic scale, it varied from about -6(which means vary strong evidence for the CAR model) to 6 (which means very strong evidence for the Poisson model).

DOI：10.5963/LSMR0304003

To sum up, the sensitivity of Bayes factor gives a clear idea about comparison of two nested models, since it varies significantly for a grid of parameter values, without requiring the Bayes factor to be computed for any particular class of prior distributions.

In order to demonstrate the efficacy of the previous four approaches, blinded predictions of patient survivability based upon real data have been made, and the difference between real survivability and survivability predicted from the four models is reported in Table 4 (most accurate prediction is in bold).

TABLE IV ESTIMATES FOR SURVIVAL (IN MONTHS) AND COMPARISON TO REAL DATA

Cancer's stage I and II						
		Real data	**GLM**	**MCMC**	**Cox**	**Bayes**
1-year	Survival (months)	153.00	152.69	152.35	152.65	**152.76**
	Survival (% of patients)	95.03	94.84	94.63	94.82	**94.88**
3-years	Survival (months)	135.00	134.65	**134.88**	134.70	134.58
	Survival (% of patients)	83.85	83.64	**83.77**	83.66	83.59
5-years	Survival (months)	126.00	126.88	126.46	126.47	**126.44**
	Survival (% of patients)	78.26	78.81	78.55	78.55	**78.53**
7-years	Survival (months)	94.00	93.79	93.90	**93.99**	93.54
	Survival (% of patients)	58.39	58.25	58.32	**58.38**	58.10
Cancer's stage III						
		Real data	**GLM**	**MCMC**	**Cox**	**Bayes**
1-year	Survival (months)	6.00	5.65	5.44	5.67	**5.76**
	Survival (% of patients)	60.00	56.47	54.37	56.74	**57.56**
3-years	Survival (months)	4.00	**3.87**	3.77	3.65	3.45
	Survival (% of patients)	40.00	**38.75**	37.65	36.55	34.48
5-years	Survival (months)	4.00	**3.88**	3.88	3.69	3.65
	Survival (% of patients)	40.00	**38.77**	45.00	45.00	45.00
7-years	Survival (months)	2.00	**1.88**	1.87	1.69	1.53
	Survival (% of patients)	20.00	**18.77**	18.65	16.88	15.34
Cancer's stage IV						
		Real data	**GLM**	**MCMC**	**Cox**	**Bayes**
1-year	Survival (months)	5.00	4.65	**4.80**	4.69	4.46
	Survival (% of patients)	55.56	51.72	**53.32**	52.08	49.51
3-years	Survival (months)	2.00	2.80	2.65	2.70	**2.55**
	Survival (% of patients)	22.22	31.10	29.50	29.99	**28.30**
5-years	Survival (months)	2.00	**1.85**	1.46	1.68	1.66
	Survival (% of patients)	22.22	**20.61**	16.18	18.72	18.47
7-years	Survival (months)	2.00	**1.99**	1.59	1.70	1.45
	Survival (% of patients)	22.22	**22.08**	17.63	18.88	16.14

As we can see in Table 4, all four techniques can be used and are suggested for survivability prediction; there are not any great differences between them, as regards both cancer's stage (I-II, III, and IV) and years of survival (1, 3, 5, and 7 years). General Linear Models, Markov Chain Monte Carlo methods, Cox regression and Bayesian statistics are proven to be good methods to predict survivability of lip cancer patients.

REFERENCES

[1] E. Cantoni and E. Ronchetti, E. Robust inference for generalized linear models, Journal of the American Statistical Association 96, 1022-1030, 2001.

[2] R. Huggins A robust approach to the analysis of repeated measures, Biometrics 49, 715-720, 1993.

[3] A. Richardson and A. Welsh. Robust restricted maximum likelihood in mixed linear models, Biometrics 51, 1429-1439, 1995.

[4] A. Richardson. Bounded influence estimation in mixed linear models, Journal of the American Statistical Association 92, 154-161, 1997.

[5] P. Gill. A robust mixed linear model analysis for longitudinal analysis, Statistics in Medicine 19, 975-987, 2000.

[6] K. Yau and A. Kuk. Robust estimation in generalized linear mixed models, Journal of the Royal Statistical Society Series B 64, 101-117, 2002.

[7] S. Sinha. Robust analysis of generalized linear mixed models, Journal of the American Statistical Association 99, 451-460, 2004.

DOI：10.5963/LSMR0304003

[8] M. Noh and Y. Lee. Robust modeling for inference from generalized linear model classes, Journal of the American Statistical Association 102(479), 1059-1072, 2007.

[9] S. Meyn and R. Tweedie. Stability of Markovian processes. I. Criteria for disctrete-time chains. Advances in Applied Probabilities 24(3), 542-574, 1992.

[10] M. Cowles and B. Carlin. Markov chain Monte Carlo convergence diagnostics: A comparative review, Journal of the American Statistical Association 91, 883-904, 1996.

[11] G. Fishman. Monte Carlo: Concepts, algorithms and applications, Springer-Verlag, New York, 1996.

[12] C. Geyer. Practical Markov chain Monte Carlo, Statistical Science 7, 473-511, 1992.

[13] G. Jones and J. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo, Statistical Science 16,: 312-334, 2001.

[14] H. Damerdji. Strong consistency and other properties of the spectral variance estimator, Management Science 37, 1424-1440, 1994.

[15] S. Lipsitz and J. Ibrahim. A conditional model for incomplete covariates in parametric regression models, Biometrika 83, 916-922, 1996a.

[16] A. Herring and J. Ibrahim. Likelihood-based methods for missing covariates in the Cox proportional hazards model, Journal of the American Statistical Association 96(453), 292-302, 2001.

[17] M. Schluchter and K. Jackson. Log-linear analysis of censred survival data with partially observed covariates, Journal of the American Statistical Association 84,42-52, 1989.

[18] D. Lin and Z. Ying. Cox regression with incomplete covariate measurements, Journal of the American Statistical Association 88, 1341-1349, 1993.

[19] M. Pugh, J. Robins, S. Lipsitz and D. Harrington. Inference in the Cox proportional hazards model with missing covariate data, Technical Report, Dana-Farber Cancer Institute, Division of Biostatistical Science, Boston, 1993.

[20] M. Reilly and M. Pepe. A mean score method for missing and auxiliary covariate data in regression models, Biometrika 82, 299-314, 1995.

[21] H. Zhou and M. Pepe. Auxiliary covariate data in failure time regression, Biometrika 82, 139-149, 1995.

[22] S. Lipsitz and J. Ibrahim. Using the EM algorithm for survival data with incomplete categorical covariats, Lifetime Data Analysis 2, 5-14, 1996b.

[23] M. Paik. Multiple imputation for the Cox proportional hazards model with missing covariates, Lifetime Data Analysis 3, 289-298, 1997.

[24] H. Chen and R. Little. Proportional hazards regression with missing covariates, Journal of the American Statistical Association 94, 896-908, 1999.

[25] R. Kass and A. Raftery. Bayes factors, Journal of the American Statistical Association 90,773-795, 1995.

[26] S. Sinharay and H. Stern. On the sensitivity of Bayes factors to the prior distribution, The American Statistician 56(3), 196-201, 2002.

[27] A. Ntomouchtsis, G. Koloutsos, N. Kechagias, K. Kitikidou, C. Tsompanidou, M. Lazaridou, C. Andreadis, K. Vachtsevanos and K. Andoniades. Carcinoma of the lip. A 10-year retrospective analysis, Proceedings of the 19th Congress of the European Association for Cranio Maxillo Facial Surgery, Bologna Italy, 9-12 September 2008. Oral Session No O. 478, 2008.

[28] N. Breslow and D. Clayton. Approximate inference in generalized linear mixed models, Journal of the American Statistical Association 88, 9-25, 1993.

[29] M. Haran and L. Tierney. Perfect sampling for a Bayesian spatial model, Technical Report, Pennsylvania State University, Dept. of Statistics, 2004.

[30] G. Jones, M. Haran, B. Caffo and R. Neath. Fixed-width output analysis for Markov chain Monte Carlo, Journal of the American Statistical Association 101(476),1537-1547, 2006.

[31] N. Cressie, H. Stern and D. Wright. Mapping rates associated with polygons, Journal of Geographical Systems 2, 61-69, 2000.

[32] H. Stern and N. Cressie. Bayesian and constrained Bayesian inference for extremes in epidemiology, Proceedings of the Epidemiology Section, American Statistical Association, pp.11-20, 1995.

DOI：10.5963/LSMR0304003