COMPARISON OF COX REGRESSION AND PARAMETRIC MODELS FOR SURVIVAL ANALYSIS OF GENETIC VARIANTS IN HNF1B GENE RELATED TO AGE AT ONSET OF CANCER

Kesheng Wang1,*, Xuefeng Liu2, Yue Pan3, Daniel Owusu1, Chun Xu4
1Department of Biostatistics and Epidemiology, East Tennessee State University, Johnson City, TN 37614, USA
2Department of Systems Leadership and Effectiveness Science, University of Michigan, Ann Arbor, MI 48109-5482, USA
3Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL 33136, USA
4Department of Health and Biomedical Sciences, University of Texas Rio Grande Valley, Brownsville, TX 78520, USA

Abstract: Semi-parametric Cox regression and parametric methods have been used to analyze survival data of cancer; however, no study has focused on the comparison of survival models in genetic association analysis of age at onset (AAO) of cancer. The Hepatocyte nuclear factor-1-beta (HNF1B) gene has been associated with risk of endometrial and prostate cancers; however, no study has focused on the effect of HNF1B gene on the AAO of cancer. This study examined 23 single nucleotide polymorphisms (SNPs) within the HNF1B gene in the Marshfield sample with 716 cancer cases and 2,848 non-cancer controls. Cox proportional hazards models in PROC PHREG and parametric survival models (including exponential, Weibull, log-normal, log-logistic, and gamma models) in PROC LIFEREG in SAS 9.4 were used to detect the genetic association of HNF1B gene with the AAO. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare the Cox models and parametric survival models. Both AIC and BIC values showed that the Weibull distribution is the best model for all the 23 SNPs and the Gamma distribution is the second best. The top two SNPs are rs4239217 and rs7501939 with time ratio (TR) =1.08 (p<0.0001 for the AA and AG genotypes, respectively) and 1.07 (p=0.0004 and 0.0002 for CC and CT genotypes, respectively) based on the Weibull model, respectively. This study shows that the parametric Weibull distribution is the best model for the genetic association of AAO of cancer and provides the first evidence of several genetic variants within the HNF1B gene associated with AAO of cancer.

Key words: Cancer, age at onset, HNF1B, SNP, survival analysis, Cox regression, parametric models

1. Introduction

Survival analysis methods, including non-parametric Kaplan-Meier method (including log-rank test and Wilcoxon test), semi-parametric Cox proportional hazards model as well as parametric methods (such as exponential, Weibull, gamma, log-normal, and log-logistic models) have been used in cancer survival studies. However, previous studies have shown inconsistent results for survival analysis methods in cancer survival studies. For example, the Cox model is similar to the exponential model (Pourhoseingholi et al., 2007); while the Weibull and exponential models were similarly the best models in the survival analysis of stomach cancer.
(Moghimi-Dehkordi et al., 2008). In other studies, the Weibull model was shown to be better than the exponential model in survival analysis of patients with gastric cancer in Iran (Baghestani et al., 2009), and was the best model for patients with gastric cancer (Zhu et al., 2011) and colorectal cancer in Iran (Baghestani et al., 2015). The log-normal survival model may have a good fit for the gallbladder cancer dataset (Wang et al., 2010); while the log-logistic model with gamma frailty is the better-fit model in the survival analysis of gastrointestinal cancer in northern Iran (Ghadimi et al., 2011). Another study has shown parametric models such as Weibull model, log-normal and gamma models may perform better than Cox models in identifying prognostic factors for oral cancer (Köhler and Kowalski, 2012).

Family and twin studies have indicated that genetic and environmental factors and their interactions contribute to the development of human cancers, with a heritability of 22% for ovarian cancer, 35% for colorectal cancer (Lichtenstein et al., 2000), 25-30% for breast cancer (Lichtenstein et al., 2000; Czene et al., 2002; Locatelli et al., 2004), and 42-58% for prostate cancer (Lichtenstein et al., 2000; Hjelmborg et al., 2014). The Hepatocyte nuclear factor-1-beta (HNF1B) gene (also known as FJHN; HNF2; TCF2; HPC11; LF-B3; MODY5; TCF-2; VHNF1; HNF-1B; HNF1beta) is located at 17q12 (Abbott et al., 1990; Bach et al., 1991; Gudmundsson et al., 2007) and a member of the homeodomain-containing superfamily of transcription factors (Bach et al., 1991). HNF1B may play a role in renal cell carcinoma (Rebouissou et al., 2005), ovarian cancer, as well as gastric, pancreatic, and colorectal cell lines (Terasawa et al., 2006). Recently, several candidate gene and genome-wide association studies (GWAS) reported that several single nucleotide polymorphisms (SNPs such as rs7501939 and rs4430796) in the HNF1B gene were associated with the risks of endometrial and prostate cancers (Gudmundsson et al., 2007; Spurdle et al., 2011; Wang et al., 2014). Age at onset (AAO) of complex diseases such as cancers have also genetic components (e.g., Claus., 1990; Shugart et al., 2000; Pankraz et al., 2005; Reeves et al., 2009; Chen et al., 2013); however, no study has examined the effect of HNF1B gene on AAO of cancer. The semi-parametric Cox proportional hazards model has been used to examine the associations of genetic variants with AAO of cancer. For example, previous studies tested the association between p53 and DNMT3b polymorphisms and AAO of colorectal cancer using the Cox proportional hazards regression model (Jones et al., 2004; Krüger et al., 2005; Sotamaa et al., 2005; Jones et al., 2006). Furthermore, Kulminski et al. (2011) reported the association between apolipoprotein E (APOE) e2/3/4 polymorphism and AAO of cancer (all sites but skin) using data on 3924 participants of the Framingham Heart Study Offspring cohort in Cox regression model; while Chen et al. (2013) examined the association of 1456 SNPs in 128 cell cycle-related genes and 31 DNA repair-related genes in 485 non-Hispanic Whites to determine whether there were SNPs associated with AAO of colorectal cancer. Recently, Wang et al. (2015) examined the associations of 220 SNPs within the PTPRN2 gene with the AAO of cancer using the Cox regression. However, no study has compared the survival models in genetic association analysis of age at onset (AAO) of cancer.

This study was to identify the best model by comparing the Cox proportional hazards models and parametric survival models in genetic association analysis of the HNF1B gene with the AAO of cancer in a Caucasian sample.

2. Subjects and Methods

2.1. The Marshfield sample

The Marshfield sample was selected from the publicly available data in A Genome-Wide Association Study on Cataract and HDL in the Personalized Medicine Research Project Cohort - Study Accession: phs000170.v1.p1 (dbGaP). Details about the participants were described elsewhere (McCarty et al., 2005, 2008). Cases were defined as any diagnosed cancer excluding minor skin cancer.
and AAO of cancer was defined by the date of the earliest cancer diagnosis in the registry. Covariates included in this study were age, gender, alcohol use in the past month (yes or no), obesity status, and smoking status (never smoking, current smoking and past smoking). Obesity was defined as a body mass index (BMI) ≥ 30 kg/m². Genotyping data using the ILLUMINA Human660W-Quad_v1_A were available for 3894 individuals. The genotypes of 23 SNPs within the HNF1B gene were available in this data.

2.2. Descriptive Statistics and Quality Control

Categorical variables were presented as frequencies and percentages, while continuous variables were reported as the means ± standard deviation (SD). HelixTree Software (http://www.goldenhelix.com/SNP_Variation/HelixTree/index.html) was used to assess control genotype data for conformity with Hardy-Weinberg equilibrium (HWE). Genotype call rates and minor allele frequency (MAF) were also calculated. To account for population stratification, the principal-component analysis approach (Price et al., 2006) in HelixTree software was used to identify outlier individuals (Wang et al., 2012). Based on the principal components analysis of the first 5 principal components using HelixTree and genome-wide genotype data, we removed outlier individuals. Consequently, 3564 Caucasian individuals were included in the analysis (716 cancer cases and 2848 controls).

2.3. Cox Proportional Hazards Model

The proportional hazards model or Cox regression model (Cox, 1972) is widely used in the analysis of time-to-event data to explain the effect of explanatory variables on hazard rates (Cantor, 2007; George et al., 2014).

\[ h(t|x) = h_0(t) \exp(\beta_1 x_1 + \ldots + \beta_p x_p) \]  

where \( h(t|x) \) is the hazard function at time t for a subject with a set of predictors \( x_1, \ldots, x_p \), \( h_0(t) \) is the baseline hazard function, and \( \beta_1, \ldots, \beta_p \) are the model parameters describing the effect of the predictors on the overall hazard. Then the hazard ratio (HR) is defined as the ratio of predicated hazard rates under two different values of a predictor variable (George et al., 2014). The PHREG procedure in SAS was used to fit the Cox regression model by considering censoring and maximizing the partial likelihood function.

2.4. Parametric Survival Models

Some commonly assumed parametric distributions in survival models include exponential, Weibull, gamma, log-normal, and log-logistic (Klein and Moeschberger, 2003).

\[ \ln(T) = \beta_1 x_1 + \ldots + \beta_p x_p + \ln(\varepsilon) \]  

where \( T \) is the time to event; \( x_1, \ldots, x_p \), and \( \beta_1, \ldots, \beta_p \) are predictor variables and their corresponding coefficients, respectively; \( \varepsilon \) is the error term assumed to have a particular parametric distribution; and \( \ln(\varepsilon) \) is the natural log of the error term (George et al., 2014). The exponentials of the \( \beta \) coefficients may be interpreted as the time ratio (TR) (Hernán et al., 2005; George et al., 2014; Kasza et al., 2014). If TR >1, the event is less likely to occur as it means it will take longer for the event to happen; whereas if TR <1, the event is more likely to happen. The LIFEREG procedure in SAS fits parametric survival models, where the link function can be
taken from a class of distributions that include exponential, Weibull, log-normal, log-logistic, and gamma distributions.
2.5. Evaluation Criteria for Goodness of Fit

The Akaike information criterion (AIC) was used as a measure of goodness of model fit that balances model fit against model simplicity (Akaike, 1979, 1981); while the Bayesian information criterion (BIC) was used as a similar measure (Simonoff, 2003).

\[
AIC = -2\ln\{p(x|\hat{\theta})\} + 2k
\]

(3)

and

\[
BIC = -2\ln\{p(x|\hat{\theta})\} + k\ln n
\]

(4)

where \(x\) is the random variable, \(\hat{\theta}\) is the maximum likelihood estimate, \(k\) is the number of parameters, and \(n\) is the sample size. Note that model with smaller AIC and BIC values fits the data better.

2.6. Survival Analysis of Age at Onset of Cancer

The assessment of the association between genotypes of each SNP and AAO was initially performed using the log-rank test and Wilcoxon test in Kaplan–Meier (KM) survival analysis using LIFETEST procedure. The KM survival curves were used to plot the survival function. The PHREG procedure in SAS was used to fit the Cox model while the LIFEREG procedure was used to fit parametric survival models including the exponential, Weibull, log-normal, log-logistic, and gamma distributions. Multivariate Cox regression analysis and parametric survival analyses were conducted to detect associations of each SNP with AAO adjusting for gender, alcohol use in the past month, smoking status and obesity status. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to compare the Cox regression and parametric survival models. Descriptive statistics, the KM survival analysis, Cox regression, and parametric model analyses were conducted with SAS v.9.4 (SAS Institute, Cary, NC, USA). SAS codes are listed in Appendix.

3. Results

3.1. Genotype Quality Control and Descriptive Statistics

All the 23 SNPs were in HWE in the controls (p>0.001) with MAF >5%. The demographic characteristics of the subjects in the study are presented in Table 1. There were slightly more females than males in both cases and controls. Age ranged from 46 to 90 years and AAO of cancer ranged from 23 to 90 years.

3.2. Comparison of Cox Regression and Parametric Models using PROC PHREG and PROC LIFEREG

The estimates of AIC and BIC consistently showed that the Weibull distribution was the best model for all the 23 SNPs and the Gamma distribution was the second best. Table 2 shows AIC and BIC for the different models for 5 SNPs associated with AAO in the Weibull model (p<0.05). The top two SNPs were rs4239217 and rs7501939, in that order. For rs4239217, the Weibull distribution was the best model (AIC and BIC are 5582.38 and 5623.38, respectively) and the Gamma distribution was the second (AIC and BIC are 5583.94 and 5629.6, respectively). For the SNP rs7501939, the AIC and BIC values showed consistent results that the Weibull distribution was the best model (AIC and BIC are 5582.38 and 5623.38, respectively) and the Gamma model was the second best (AIC and BIC are 5583.94 and 5629.6, respectively).
3.3. Survival Analysis of Age at Onset using the Weibull Model

The results of the parametric survival analysis using the Weibull model is presented in Table 3. Five SNPs (rs3110649, rs757210, rs430796, rs4239217 and rs7501939) showed associations with AAO of cancers. The top SNP was rs4239217 with TR=1.08 and the second signal was rs7501939 with TR=1.07. The Kaplan–Meier survival curves for different genotypes of SNPs rs4239217 and rs7501939 are shown in Figures 1 and 2, respectively. For rs4239217, the mean AAO was approximately 5.5 and 4.6 years later for individuals with AA genotype and AG genotype compared to those with GG genotype (p=0.0002 based on the log-rank test and p<0.0001 based on the Wilcoxon test). For rs7501939, the mean AAO was approximately 5.2 and 4.3 years later for individuals with CC genotype and CT genotype compared to those with TT genotype (p=0.0058 based on the log-rank test and p=0.0002 based on Wilcoxon test).

4. Discussion

In the present study, we explored the associations of 23 HNF1B SNPs with the AAO of cancer. Using the AIC and BIC, the Weibull distribution was found to be the best model for genetic association of polymorphisms within the HNF1B gene with the AAO of cancer. To our knowledge, this is the first study to compare the Cox regression and parametric survival models in genetic association analysis of AAO of cancer. It is also the first candidate gene study to provide evidence of several genetic variants (rs3110649, rs757210, rs430796, rs4239217 and rs7501939) within the HNF1B gene which may be involved in the AAO of cancer.

Semi-parametric Cox regression and parametric survival methods (such as exponential, Weibull, gamma, log-normal, and log-logistic models) have been used in cancer survival studies. However, previous studies have shown inconsistent results for survival analyses in cancers. For example, one study showed that the Cox regression is similar to the exponential model (Pourhoseingholi et al., 2007); while another study showed parametric models such as Weibull model, lognormal and gamma models may perform better than Cox model in oral cancer (Köhler and Kowalski, 2012). Furthermore, some studies favored the Weibull model in stomach cancer (Moghimi-Dehkordi et al., 2008), gastric cancer (Baghestani et al., 2009; Zhu et al., 2011) and colorectal cancer (Baghestani et al., 2015). In addition, one study found that the log-normal survival model may have a good fit for the gallbladder cancer (Wang et al., 2010); while the log-logistic model with gamma frailty is the best model in gastrointestinal cancer in northern Iran (Ghadimi et al., 2011). Several studies have used the semi-parametric Cox model to examine the association of genetic variants with AAO of colorectal cancer (Jones et al., 2004; Krüger et al., 2005; Sotamaa et al., 2005; Jones et al., 2006; Chen et al., 2013) and of cancer (all sites but skin) (Kulminski et al., 2011; Wang et al., 2015). However, no study was found to have examined associations between genetic variants and AAO of cancer using parametric models (including exponential, Weibull, log-normal, log-logistic and gamma models). Therefore, the current study is the first attempt to compare the Cox regression with parametric survival models in genetic association analysis of AAO of cancer. Furthermore, in consistent with some previous studies (such as Moghimi-Dehkordi et al., 2008; Baghestani et al., 2009; Zhu et al., 2011; Baghestani et al., 2015), our results showed that the Weibull distribution is the best model for genetic associations of all 23 SNPs within the HNF1B gene with AAO of cancer. Moreover, we found that the Gamma distribution is the second best model. The differences of these comparisons may be due to different cancer types, sample size, and sample origins. On the other hand, we performed genetic association study of the AAO of cancer using parametric survival models.

It was shown that HNF1B regulated the expression of polycystic kidney and hepatic disease-1 (PKHD1) and therefore HNF1B may function as a tumor suppressor gene in chromophobe renal cell carcinogenesis (Rebouissou et al., 2005); while HNF1B may be involved in the development of ovarian cancers, gastric, pancreatic, and colorectal cell lines (Terasawa et al., 2006). Recently, several candidate genes and GWAS...
studies observed that rs4430796 within HNF1B gene was associated with prostate cancers (Gudmundsson et al., 2007; Eeles et al., 2008; Levin et al., 2008; Thomas et al., 2008; Waters et al., 2009; Berndt et al., 2011; Spurdle et al., 2011, Wang et al., 2014) and endometrial cancer (Spurdle et al., 2011; Wang et al., 2014). The
signal of rs7501939 was reported to be associated with prostate cancers (Eeles et al., 2008; Levin et al., 2008; Wang et al., 2014) and endometrial cancer (Setiawan et al., 2012; Wang et al., 2014). Several studies also found that rs4430796 was associated with lung cancer in Chinese population (Sun et al., 2011), and prostate cancer in Korean men (Kim et al., 2008), Chinese men (Zhang et al., 2012), and African American men (Chornokur et al., 2013). However, no study has focused on the effect of HNF1B gene on AAO of cancer. In the present study, we provided the first evidence that two previously cancer risk associated SNPs (rs430796 and rs7501939) within HNF1B gene were associated with AAO of cancer. We added that 3 more SNPs (rs3110649, rs757210, and rs4239217) were associated with AAO of cancer.

Studies suggest that type 2 diabetes (T2D) might share the same genetic link to prostate cancer. HNF1A S319 was indicated to be associated with earlier AAO of T2D in women (Hegele et al., 2000). Two SNPs (rs7501939 and rs4430796) of HNF1B were reported to be associated with T2D in Chinese as well as in Caucasians (Gudmundsson et al., 2007). Recently, the results of some GWAS studies provided support for a shared genetic contribution to the risk of T2D and prostate cancer. For example, in the study by Gudmundsson et al. (2007), the A allele of rs4430796 and C allele of rs7501939 variants in HNF1B/TCF2 showed positive associations with prostate cancer (OR>1.0) but was protective against T2D (OR<1.0). A later study confirmed the association of rs4430796 with T2D and prostate cancer and suggested that T2D had a protective effect on prostate cancer risk (Piece et al., 2010). A meta-analysis examined the two variants (rs4430796 and rs7501939) and found they had pleiotropic effects on T2D and prostate cancer (Elliott et al., 2010). The present study added that the HNF1B gene may play a role in the development of cancer.

There are some limitations in this study. First, the definition of cancer status in the Marshfield sample was broad (including any diagnosed cancer omitting minor skin cancer) which may result in genetic and phenotypic heterogeneity into the genetic association analysis. It would be more informative to investigate the association of HNF1B with specific types of cancer. Second, our current findings might be subject to type I error and findings need to be replicated in additional samples. In addition, in the present study, we just used the original Cox regression model and parametric survival models. Interestingly, investigators have extended original Cox regression and parametric models. For example, Li et al. (2016) recently developed the proportional generalized odds (PGO) model, which covers the proportional odds (PO) model (Bennett, 1983; Pettitt. 1984) and the generalized proportional odds (GPO) model (Dabrowska and Doksum, 1988). On the other hand, Musrafa et al. (2016) proposed a new four parameters Weibull model called the Weibull Generalized Flexible Weibull extension (WGFWE) distribution and Alkarni (2016) introduced a new family of models for lifetime data called generalized extended Weibull power series family of distributions by compounding generalized extended Weibull distributions and power series distributions; while Pu et al. (2016) proposed a new class of five parameters gamma-exponentiated or generalized modified Weibull (GEMW) distribution. In the present study, we just tested two parameters Weibull model. In the future, it will be prospective to test and apply these extended survival models in the genetic association of AAO of complex diseases.

There are also several strengths in this study. First, our sample size was relatively large for this type of study. Second, we compared the semi-parametric Cox regression and parametric models in the genetic association of AAO of cancer. Third, we examined 23 SNPs within the HNF1B gene and especially identified 2 cancer and T2D associated SNPs (rs4430796 and rs7501939) influencing the AAO of cancer.

In conclusion, the results demonstrate that the parametric Weibull model performed better than Cox regression and other parametric models (including exponential, log-normal, log-logistic and gamma models) for the genetic association of AAO of cancer. Furthermore, this study provides evidence of several genetic variants within the HNF1B gene influencing AAO of cancer. These findings may serve as a resource for replication in other populations. Future functional study of this gene may help to better characterize the genetic architecture of the AAO of cancer.
Acknowledgement

Funding support for the Personalized Medicine Research Project (PMRP) was provided through a cooperative agreement (U01HG004608) with the National Human Genome Research Institute (NHGRI), with additional funding from the National Institute for General Medical Sciences (NIGMS). The samples used for PMRP analyses were obtained with funding from Marshfield Clinic, Health Resources Service Administration Office of Rural Health Policy grant number D1A RH00025, and Wisconsin Department of Commerce Technology Development Fund contract number TDF FYO10718. Funding support for genotyping, which was performed at Johns Hopkins University, was provided by the NIH (U01HG004438). Assistance with phenotype harmonization and genotype cleaning was provided by the eMERGE Administrative Coordinating Center (U01HG004603) and the National Center for Biotechnology Information (NCBI). The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000170.v1.p1. This study was approved by the Internal Review Board (IRB), East Tennessee State University.

Role of the funding sources

No founding source is given for the present paper.

Conflict of interest

All authors have reported no financial interests or potential conflicts of interest.

References


430 COMPARISON OF COX REGRESSION AND PARAMETRIC MODELS FOR SURVIVAL ANALYSIS OF GENETIC VARIANTS IN HNF1B GENE RELATED TO AGE AT ONSET OF CANCER


434 COMPARISON OF COX REGRESSION AND PARAMETRIC MODELS FOR SURVIVAL ANALYSIS OF
GENETIC VARIANTS IN HNF1B GENE RELATED TO AGE AT ONSET OF CANCER


Table 1: Descriptive Characteristics of Cases and Controls in the Study Sample

<table>
<thead>
<tr>
<th></th>
<th>Non-Cancer</th>
<th>Cancer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>2848</td>
<td>716</td>
</tr>
<tr>
<td>Sex, N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td>1135(40%)</td>
<td>340(47%)</td>
</tr>
<tr>
<td>Females</td>
<td>1713(60%)</td>
<td>376(53%)</td>
</tr>
<tr>
<td>Obesity, N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1700(60%)</td>
<td>422(59%)</td>
</tr>
<tr>
<td>Yes</td>
<td>1148(40%)</td>
<td>294(41%)</td>
</tr>
<tr>
<td>Alcohol use, N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1060(37%)</td>
<td>288(40%)</td>
</tr>
<tr>
<td>Yes</td>
<td>1783(63%)</td>
<td>425(60%)</td>
</tr>
<tr>
<td>Smoking status, N (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1487(52%)</td>
<td>327(46%)</td>
</tr>
<tr>
<td>Current</td>
<td>254(9%)</td>
<td>54(7%)</td>
</tr>
<tr>
<td>Past</td>
<td>1104(39%)</td>
<td>331(47%)</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>65.1±11.3</td>
<td>71.1±10.3</td>
</tr>
<tr>
<td>Range</td>
<td>46-90</td>
<td>46-90</td>
</tr>
<tr>
<td>Age at onset, years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± SD</td>
<td>-</td>
<td>64.2±12.8</td>
</tr>
<tr>
<td>Range</td>
<td>-</td>
<td>23-90</td>
</tr>
</tbody>
</table>
Table 2: Results of the Cox Regression and Parametric Models in the Multivariate Survival Analysis of AAO of Cancer

<table>
<thead>
<tr>
<th>Models</th>
<th>AIC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>BIC&lt;sup&gt;b&lt;/sup&gt;</th>
<th>AIC&lt;sup&gt;c&lt;/sup&gt;</th>
<th>BIC&lt;sup&gt;d&lt;/sup&gt;</th>
<th>AIC&lt;sup&gt;e&lt;/sup&gt;</th>
<th>BIC&lt;sup&gt;f&lt;/sup&gt;</th>
<th>AIC&lt;sup&gt;g&lt;/sup&gt;</th>
<th>BIC&lt;sup&gt;h&lt;/sup&gt;</th>
<th>AIC&lt;sup&gt;i&lt;/sup&gt;</th>
<th>BIC&lt;sup&gt;j&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox</td>
<td>7955.52</td>
<td>7987.47</td>
<td>7952.80</td>
<td>7984.75</td>
<td>7917.22</td>
<td>7949.15</td>
<td>7943.91</td>
<td>7975.87</td>
<td>7935.73</td>
<td>7967.67</td>
</tr>
<tr>
<td>Weibull</td>
<td>5594.59</td>
<td>5635.67</td>
<td>5590.97</td>
<td>5632.06</td>
<td>5569.69</td>
<td>5610.73</td>
<td>5582.38</td>
<td>5623.38</td>
<td>5578.04</td>
<td>5619.11</td>
</tr>
<tr>
<td>Exponential</td>
<td>7344.54</td>
<td>7381.07</td>
<td>7344.32</td>
<td>7380.84</td>
<td>7314.77</td>
<td>7351.26</td>
<td>7343.91</td>
<td>7380.43</td>
<td>7343.91</td>
<td>7380.43</td>
</tr>
<tr>
<td>Log-logistic</td>
<td>5685.02</td>
<td>5726.11</td>
<td>5678.06</td>
<td>5719.15</td>
<td>5654.21</td>
<td>5695.25</td>
<td>5665.00</td>
<td>5706.09</td>
<td>5661.89</td>
<td>5702.97</td>
</tr>
<tr>
<td>Log-normal</td>
<td>5705.60</td>
<td>5746.69</td>
<td>5700.06</td>
<td>5741.15</td>
<td>5676.62</td>
<td>5717.67</td>
<td>5690.32</td>
<td>5731.40</td>
<td>5686.24</td>
<td>5727.32</td>
</tr>
<tr>
<td>Gamma</td>
<td>5595.50</td>
<td>5641.16</td>
<td>5592.22</td>
<td>5637.88</td>
<td>5571.02</td>
<td>5616.63</td>
<td>5583.94</td>
<td>5629.60</td>
<td>5579.54</td>
<td>5625.18</td>
</tr>
</tbody>
</table>

<sup>a</sup> AIC and BIC for rs3110649 adjusted for sex, alcohol use, smoking status, and obesity; <sup>b</sup> AIC and BIC for rs757210 adjusted for sex, alcohol use, smoking status, and obesity; <sup>c</sup> AIC and BIC for rs430796 adjusted for sex, alcohol use, smoking status, and obesity; <sup>d</sup> AIC and BIC for rs4239217 adjusted for sex, alcohol use, smoking status, and obesity; <sup>e</sup> AIC and BIC for rs7501939 adjusted for sex, alcohol use, smoking status, and obesity.
### Table 3: Survival Analysis of the 5 SNPs Associated with AAO Using the Weibull Model

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position (bp)</th>
<th>Allele</th>
<th>MA F&lt;sup&gt;b&lt;/sup&gt;</th>
<th>HWE&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Genotype</th>
<th>β&lt;sup&gt;d&lt;/sup&gt;</th>
<th>95% CI&lt;sup&gt;e&lt;/sup&gt;</th>
<th>p-value&lt;sup&gt;f&lt;/sup&gt;</th>
<th>TR&lt;sup&gt;g&lt;/sup&gt;</th>
<th>95% CI&lt;sup&gt;h&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3110649</td>
<td>33144293</td>
<td>T</td>
<td>0.21</td>
<td>0.374</td>
<td></td>
<td>0.063</td>
<td>0.003,0.12</td>
<td>0.0387</td>
<td>1.07</td>
<td>1.01,1.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>0.063</td>
<td>0.003,0.12</td>
<td>0.0387</td>
<td>1.07</td>
<td>1.01,1.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT</td>
<td>0.045</td>
<td>-0.02,-0.10</td>
<td>0.143</td>
<td>1.05</td>
<td>0.98,1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TT</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>rs757210</td>
<td>33170628</td>
<td>A</td>
<td>0.37</td>
<td>0.613</td>
<td></td>
<td>-0.056</td>
<td>-0.097,-0.019</td>
<td>0.0039</td>
<td>0.93</td>
<td>0.89,0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>-0.056</td>
<td>-0.097,-0.019</td>
<td>0.0039</td>
<td>0.93</td>
<td>0.89,0.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>-0.001</td>
<td>-0.028,0.025</td>
<td>0.933</td>
<td>1.00</td>
<td>0.97,1.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>rs430796</td>
<td>33172153</td>
<td>G</td>
<td>0.48</td>
<td>0.129</td>
<td></td>
<td>0.034</td>
<td>-0.001,0.069</td>
<td>0.0567</td>
<td>1.03</td>
<td>1.00,1.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.034</td>
<td>-0.001,0.069</td>
<td>0.0567</td>
<td>1.03</td>
<td>1.00,1.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>0.034</td>
<td>0.002,0.066</td>
<td>0.0372</td>
<td>1.03</td>
<td>1.01,1.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>rs4239217</td>
<td>33173100</td>
<td>G</td>
<td>0.4</td>
<td>0.18</td>
<td></td>
<td>0.075</td>
<td>0.038,0.111</td>
<td>&lt;0.000&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1.08</td>
<td>1.04,1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AA</td>
<td>0.075</td>
<td>0.038,0.111</td>
<td>&lt;0.000&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1.08</td>
<td>1.04,1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>AG</td>
<td>0.076</td>
<td>0.039,0.111</td>
<td>&lt;0.000&lt;sup&gt;1&lt;/sup&gt;</td>
<td>1.08</td>
<td>1.04,1.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GG</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>rs7501939</td>
<td>33175269</td>
<td>T</td>
<td>0.39</td>
<td>0.073</td>
<td></td>
<td>0.069</td>
<td>0.031,0.106</td>
<td>0.0004</td>
<td>1.07</td>
<td>1.03,1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CC</td>
<td>0.069</td>
<td>0.031,0.106</td>
<td>0.0004</td>
<td>1.07</td>
<td>1.03,1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CT</td>
<td>0.069</td>
<td>0.033,0.106</td>
<td>0.0002</td>
<td>1.07</td>
<td>1.03,1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TT</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
a Minor allele; b Minor allele frequency; c Hardy-Weinberg equilibrium test p-value; d Regression coefficient for AAO of cancer based on the Weibull model; e 95%CI of regression coefficient for AAO of cancer based on the Weibull model; f p-value for AAO of cancer based on Weibull model; g Time ratio (TR) for the genotype comparing with reference; h 95%CI of TR for the genotype comparing with reference.

Figure 1: Survival Function by Three Genotypes of rs4239217
Comparison of Cox regression and Parametric Models for Survival Analysis of Genetic Variants in HNF1B gene Related to Age at Onset of Cancer

APPENDIX

The following program using PROC PHREG showed one SNP rs4239217, sex, alcohol use, smoking status, and obesity with the AAO of cancer. The rs4239217 has 3 genotypes - A_A, A_G and G_G, respectively; while the G_G genotype was considered as the reference.

```plaintext
proc phreg data= canceraao;
  class sex(ref="1") obesity(ref="1") alcohol(ref="1") smoking(ref="1") rs4239217 (ref = "G_G")/param=ref;
  model canceraao*statuscan(0) = sex obesity alcohol smoking rs4239217 / risklimits ;
```

Figure 2: Survival Function by Three Genotypes of rs7501939
hazardratio rs4239217;

run;

The following program using PROC LIFEREG showed one SNP rs4239217, sex, alcohol use, smoking status, and obesity with the AAO of cancer for the Weibull distribution. Other parametric models can be tested by changing the dist (distribution) option in the model.

proc lifereg data= canceraao;

class sex obesity alcohol smoking rs4239217;

model canceraao*statuscan(0) = sex obesity alcohol smoking rs4239217 / dist=WEIBULL;

run;
Kesheng Wang  
Department of Biostatistics and Epidemiology  
College of Public Health  
East Tennessee State University, Johnson City, TN 37614, USA  
wangk@etsu.edu

Xuefeng Liu  
Department of Systems Leadership and Effectiveness Science  
School of Nursing  
University of Michigan, Ann Arbor, MI 48109-5482, USA

Yue Pan  
Department of Public Health Sciences  
Miller School of Medicine  
University of Miami, Miami, FL 33136, USA

Daniel Owusu  
Department of Biostatistics and Epidemiology  
College of Public Health  
East Tennessee State University, Johnson City, TN 37614, USA

Chun Xu  
Department of Health and Biomedical Sciences  
College of Health Affairs  
University of Texas Rio Grande Valley, Brownsville, TX 78520, USA