

Quantifying Maternal and Paternal Disease History Using Log-Rank Score with an Application to a National Cohort Study

Rui Feng^{1,*}, Hersh Patel² and George Howard³

¹*Department of Biostatistics and Epidemiology, University of Pennsylvania, USA*

²*Department of Biology, University of Pennsylvania, USA*

³*Department of Biostatistics, University of Alabama at Birmingham, USA*

Abstract: Both maternal and paternal disease history can be important predictors of the risk of common conditions such as heart disease or cancer because of shared environmental and genetic risk factors. Sometimes maternal and paternal history can have remarkably different effects on offspring's status. The results are often affected by how the maternal and paternal disease histories are quantified. We proposed using the log-rank score (LRS) to investigate the separate effect of maternal and paternal history on diseases, which takes parental disease status and the age of their disease onset into account. Through simulation studies, we compared the performance of the maternal and paternal LRS with simple binary indicators under two different mechanisms of unbalanced parental effects. We applied the LRS to a national cohort study to further segregate family risks for heart diseases. We demonstrated using the LRS rather than binary indicators can improve the prediction of disease risks and better discriminate the paternal and maternal histories. In the real study, we found that the risk for stroke is closely related with maternal history but not with paternal history and that maternal and paternal disease history have similar impact on the onset of myocardial infarction.

Keywords: Family history, stroke, risk score, maternal effect, imprinting.

1. INTRODUCTION

Complex diseases including heart diseases, hypertension, and breast cancer often aggregate within families and disease history of ancestors is often correlated with the disease outcomes of descendants due to common environmental and inherited genetic factors. Interestingly, studies have found that such association may not be equal from maternal relatives as from paternal relatives. For example, maternal history plays a larger role in Alzheimer's disease, type 2 diabetes, and breast cancer compared with paternal history [1-3]; only paternal not maternal family history was associated with lumbar spine bone mineral density [4]. Most of cohort studies have not pay attention to the possible unbalanced maternal and paternal effects due to possible lack of data, lack of interest, and mostly lack of appropriate statistical method. We believe that evaluating the effect of parental disease history on diseases and segregating maternal and paternal history can provide a better understanding of the disease risk factors, improve the accuracy of risk assessment, and allow clinicians to better target on high-risk individuals. In a recent national cohort study, we quantified parental disease history and investigated their associations with heart diseases.

Historically, paternal and maternal disease histories were simply quantified using binary indicators defined

by the presence of the disease in mothers or fathers. However, incidence of many diseases is distinct between males and females and changes over age. The commonly used binary indicator unfortunately cannot capture the additional information in the specific age at onset of the affected parent.

The log-rank statistic is a key concept to compare survivals, which is a summation of the contrasts of observed versus expected failures over all observed time intervals [5]. Within each time interval, the subcomponent of the long-rank statistic, as the difference between the observed and expected numbers of failures, can serve as a standardized measure of observed extremeness and has been used to compare two groups [6]. In a previous study [7], we have proposed a stratified log-rank score (LRS) to quantify family disease history, which accounts for the age of disease onset and disease status jointly. In this study, we would focus on separate parental influences.

Possible mechanistic explanations are often sought following a claim of imbalanced influence of paternal histories. Intuitively, the lifetime environmental and cultural factors of children are more strongly linked with their mothers than with their fathers and may contribute to a large percentage of disease variation. Also, maternal effects were widely recognized to affect neonatal or neurological disorders due to genetics or prenatal exposure to tobacco, nutrition, drug, and pollution [8-13]. In addition, a pertinent genetic mechanism called imprinting, where genes are only

*Address correspondence to this author at the Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19094, USA; Tel: 215-746-4473; Fax: 215-573-1050; E-mail: ruifeng@upenn.edu

expressed when inherited from mothers or fathers, could explain the imbalance of paternal history on diseases. For instance, individuals with maternal history of Alzheimer's disease are more susceptible to neurodegenerative disease because they often carry the genes that prevent their brains from using glucose efficiently [1, 14].

In this paper, we propose to extend the use of the LRS to quantify maternal and paternal disease history, which will capture the informative risk differences in various age-of-onset and gender groups. To understand the performance of the LRS in general study, we conducted simulations with either maternal or imprinting effect. We also illustrated the advantage of using the new score by comparing it with conventional binary indicators. At last, we applied our score to a large cohort study of heart diseases. We found that only the maternal stroke history but not the paternal history was related to the stroke incidences of offspring and that the influences from parental histories on myocardial infarction (MI) were well balanced.

2. METHOD

Suppose we have information on both parents of an individual and we know whether or not they have developed a particular disease at their current age or until their death. Different from the binary indicator, which simply indicates a person's disease status, the log-rank score takes all individuals' information in a group into consideration and gives a person a relative score to indicate his/her risk of disease development. For example, if the majority of fathers in a study developed a disease at the age of 60, and there was a father whose age at onset was 40, then his score would be relatively higher. In this particular study, we are only interested in quantify parents' disease history and thus we define a fathers' group and a mothers' group. Then the score of a mother or a father is calculated separately within the corresponding group so that the relative risk of one's disease status is only compared to others with comparable characteristics. Specifically, for the fathers' group, let $T_0 = 0$ and T_1, T_2, \dots, T_r , be the r distinct time points of observed events for all fathers. Suppose during the l^{th} time period $(T_{l-1}, T_l]$, there are n_0^l observations without an observed event (censored) and n_1^l observed events, then the risk of disease for any father in the l^{th} period $(T_{l-1}, T_l]$ is

$n_1^l / \sum_{m=l}^r n_0^m + n_1^m$. Assuming that the censorship occurs randomly and that the censored times are lower bounds for true event times [15], the log-rank score for censored events up to the k^{th} time period is defined as

the negative of the sum of the risk of disease from the beginning to the k^{th} period,

$$a_0^k = -\sum_{l=1}^k \frac{n_1^l}{\sum_{m=1}^r n_0^m + n_1^m}$$

and the LRSs for observed observations are defined as

$$a_1^k = 1 + a_0^k.$$

Obviously, the magnitude of a_0^k is an increasing function of the time period highlighting a common consensus that one is more likely to get a disease when getting older. If a father has an age of onset between T_{k-1} and T_k , he is assigned a risk score a_1^k ; otherwise, he gets a score of a_0^k if his age at censor is within $(T_{k-1}, T_k]$. This procedure is repeated for the mothers' group and then every parent is assigned a score.

Here we provide a simplified example to illustrate how the LRS is calculated. Assume there are 7 unrelated individuals in the population and their fathers have event times 49, 56, 60+, 66, 70+, 71. Then $T_1=49, T_2=56, T_3=66, T_4=71, n_0^1=n_0^2=0, n_0^3=n_0^4=1, n_1^1=n_1^2=0, n_1^3=n_1^4=1$. Thus, $a_0^1 = -1/6, a_0^2 = -(1/6+1/5) = -11/30, a_0^3 = -11/30-1/4 = -37/60, a_0^4 = -37/60-1/2 = -67/60,$ and $a_1^1 = 5/6, a_1^2 = 19/30, a_1^3 = 23/60, a_1^4 = -7/60$. For an event, the corresponding LRS is 1 minus the order statistic for this observation, which orders individuals from the smallest to largest risk; The LRS for the censored observation suggests the reduced risk compared with the observed events and its value was derived by Koch (1985) based on the argument that the total individual risk of observed events should be the same as the total risk of all censored observations given the independence of censoring and risk. As in the previous example, the summation of all event scores equals $-26/15$, which balances the total risk score at two censored observations $a_0^3 + a_0^4$.

Intuitively, one can consider this LRS as the binary score offset by the age-specific risk of the disease in a population. In order for the LRSs to be meaningful and helpful to the general public as a guidance, the number of people in the study must be large, and the censorship must exhibit a random pattern [16]. A large cohort study like the one we reported in this paper can serve as a basis to generate a standardized risk chart for doctors and researchers to access disease risk for the general public.

Other factors that cause heterogeneity in parental disease risks, such as birth cohort or ancestry admixture, can be adjusted using a stratified log-rank score [7].

3. SIMULATIONS

We simulated 1000 families with 2 parents and 1 offspring. Each individual had a disease phenotype and its age of onset available, which were either random or associated with their mothers' genotypes or their own inherited maternal alleles. For the parents to have their mothers' genotypes available, we first generated the genotypes of the 4 grandparents at a diallelic autosomal locus and their offspring inherited an allele from each parent following Mendelian inheritance. One allele at the locus was a causal allele and its mean allele frequency (MAF) varied from 10% to 50%.

Under the null hypothesis that neither parent's history is related with the disease status of the offspring, the age of onset for each offspring was generated from the exponential distribution, with the scale parameter λ equal to β_0 (mean of β_0). Under the alternative hypothesis that only maternal (not paternal) history is a risk factor for disease, the age of onset of offspring was generated similarly with the scale parameter $\beta_0 - \beta_1 G_m$ where G_m was the number of causal alleles in subjects' mothers for maternal effect mode, and the maternally inherited allele in offspring for imprinting model. We fixed β_0 at 4 then the constant hazard rate at 1/4 for those who did not carry the maternal risk alleles. Then $1/(1 - \beta_1/\beta_0)$ was the hazard ratio of the event of interest due to the inheritance of maternal risk allele copy to the offspring and its magnitude indicated the strength of the association between the imprinted allele and the event. We let the hazard ratio vary from 1.5, 2, to 3 with corresponding β_1 being 4/3, 2, and 8/3. G_m was an underlying gene and could not be observed in the real cohort data. The goal of our proposed method is to quantify the maternal and paternal disease risk and to associate the maternal history with the event of interest.

Another event of censoring was generated from the exponential distribution with a parameter $\lambda\gamma/(1-\gamma)$ so that γ percent of censoring would occur. If time to censoring was less than the time to the event, a right censoring event was recorded to indicate the person had not developed the disease yet, while if the time of the event was less than that of censoring, then that age-onset was the time of disease development.

In addition to calculating our proposed LRS as described above, we also used the paternal and maternal binary indicators defined as whether the father and the mother had the event or not at any age.

We are interested in detecting a strong association between offspring's disease status and their parental disease history. The null hypothesis is that there is no correlation between offspring and their maternal or paternal history. We used a Cox regression model [17] to estimate the effect of the maternal and paternal history on the offspring events.

The association was inferred from the hazard ratio of offspring's disease due to paternal or maternal risk score change. A nominal significance level of 0.05 was used. 10,000 simulations were conducted for each MAF under the null hypothesis and 1,000 were done for each MAF and β_1 under the alternative hypothesis. The power of these statistical tests, or the proportion of events that had significant correlations, was calculated for each of the five MAF values (10%, 20%, 30%, 40%, and 50%) and the three β_1 values (1.0, 2.0, and 3.0).

The simulations were repeated for the number of families being 200 and 500 and the percentage of censoring being 10% and 40% to monitor the type-I errors and power influenced by smaller sample sizes and various percentage of censoring.

4. REAL DATA APPLICATION

We applied our method to the *REasons for Geographic and Racial Differences in Stroke* (REGARDS) study, a national cohort study of individuals over age 45 years. The recruitment began in January 2003 and was completed in October 2007. The individuals (proband) from commercial lists of residents in the 48 contiguous US states were contacted by mail and phone calls. For those who agreed to participate, demographic information, medical history, including prior diagnosis of high blood pressure, family history, and indices of cognitive function and quality of life were collected by computer-assisted telephone interview (CATI). Following the CATI, physical measures were collected during in-home examinations including height, weight, blood pressure, blood and urine samples, electrocardiogram, and an inventory of current medications. Also, a self-administered questionnaire was provided to each participant to collect information on stroke and heart attack events of their parents and up to four siblings. As of June 2006, 22,927 participants had completed in-

home visits, and 18,945 (83%) participants returned the family history questionnaires through mail. A total of 57,269 first-degree family members of 13,995 REGARDS participants had information available about the age at the onset of stroke or MI events, age at death from other diseases, or age at the time of form completion without event. Informed consent was obtained from all participants, and the study was approved by the University of Alabama's Institutional Review Board for human use.

Proband's demographic characteristics were summarized using means and proportions. For their parents, the 10-year incidence rates for both stroke and MI were calculated for separate gender and age groups to understand the gender and age differences in the incidence.

For all proband, we calculated the LRS for their mothers and fathers separately. As a preliminary step to investigate the association between the disease event and parental history, we inspected the Kaplan-Meier survival curves of stroke and MI for different groups classified by their parental LRSs. Also, the comparison was done by using the groups with and without parental disease history.

To adjust for other known risk factors and potential confounders, we used a Cox regression model [17] to predict the risk of an event at the current age for each person in the test set, with age, gender, race, paternal or maternal history score, cholesterol level, diastolic blood pressure (DBP), high-density lipoproteins (HDL), C-reactive protein (Crp), hypertension, and diabetes status as potential predictors. A base Cox model was chosen to keep all the significant predictors with $p < 0.05$. To determine whether parental history predicted the events, an additional maternal or paternal LRS was added to the base model. For comparison, we also used the binary indicator in place of LRS. We

repeated the procedure for the two events of interest, MI and stroke, separately. Proportional hazard assumptions were checked using a Chi-square test and none of the predictors had significant violation of the assumption. The final models include all significant risk factors and parental health history information. To measure how good all models predict, we calculated the concordance index (C-index) [18, 19] using R package survAUC [20].

5. RESULTS

Type I Error of Using Family Risk Scores Under the Null Hypothesis

When $\beta_1 = 0$ or the hazard ratio equals 1, the time to event for each individual does not depend on the risk of their mothers and the maternal LRS is not associated with offspring disease status. Under this null hypothesis, we generated data with five different MAFs.

Based on 10,000 replications for each family size, the empirical type I errors of finding significant associations between maternal history and offspring's disease status by using maternal LRS or binary indicator are listed in Table 1. Under the null hypothesis of no true family risk, the empirical type I errors for both our LRS and the binary indicators were conservative at the nominal level of 0.05 (95% C.I. 0.044--0.056). The type I errors for association between paternal history and offspring's disease status were similar as expected and are not shown.

Power

The powers to detect a significant association between offspring's risk and maternal history score at 5% significance level are shown in Figure 1 for different MAFs and β from the paternal imprinting and maternal effect models. (a), (c), and (e) are from the imprinting

Table 1: Type I Errors of Finding Significant Associations Between the Offspring's Status and Parental History Scores

Association with	No. of families	Minor Risk Allele Frequency				
		0.1	0.2	0.3	0.4	0.5
Maternal LRS	200	0.056	0.052	0.054	0.055	0.050
	500	0.054	0.053	0.050	0.047	0.048
	1000	0.046	0.048	0.046	0.046	0.045
Maternal Event Indicator	200	0.049	0.049	0.050	0.052	0.054
	500	0.050	0.052	0.055	0.051	0.050
	1000	0.051	0.048	0.046	0.050	0.045

models and (b), (d), and (f) are from the maternal effect models.

The top panels of Figure 1 show the powers for different hazard ratios along MAF of 0.1 to 0.5 with 1000 families. Under the imprinting scenarios, the power of detecting significant maternal disease history increased from 6.0% to above 75% as the hazard ratio increased from 2 to 5 (corresponding to those with maternal causal allele vs. those without). For a fixed hazard ratio, as the casual allele frequencies increased

up to 0.2-0.3, the power increased; however, with larger allele frequencies, a larger proportion of mothers with offspring who do not carry the risk allele would also carry the grand-maternal risk allele, and thus can dilute the power to detect the actual association. Under the maternal effect scenarios, the power increased as HR or MAF increased. The power in dash lines were from the association tests using the binary maternal history indicator and were almost constantly around the type-I error rates, much lower than using our maternal LRS.

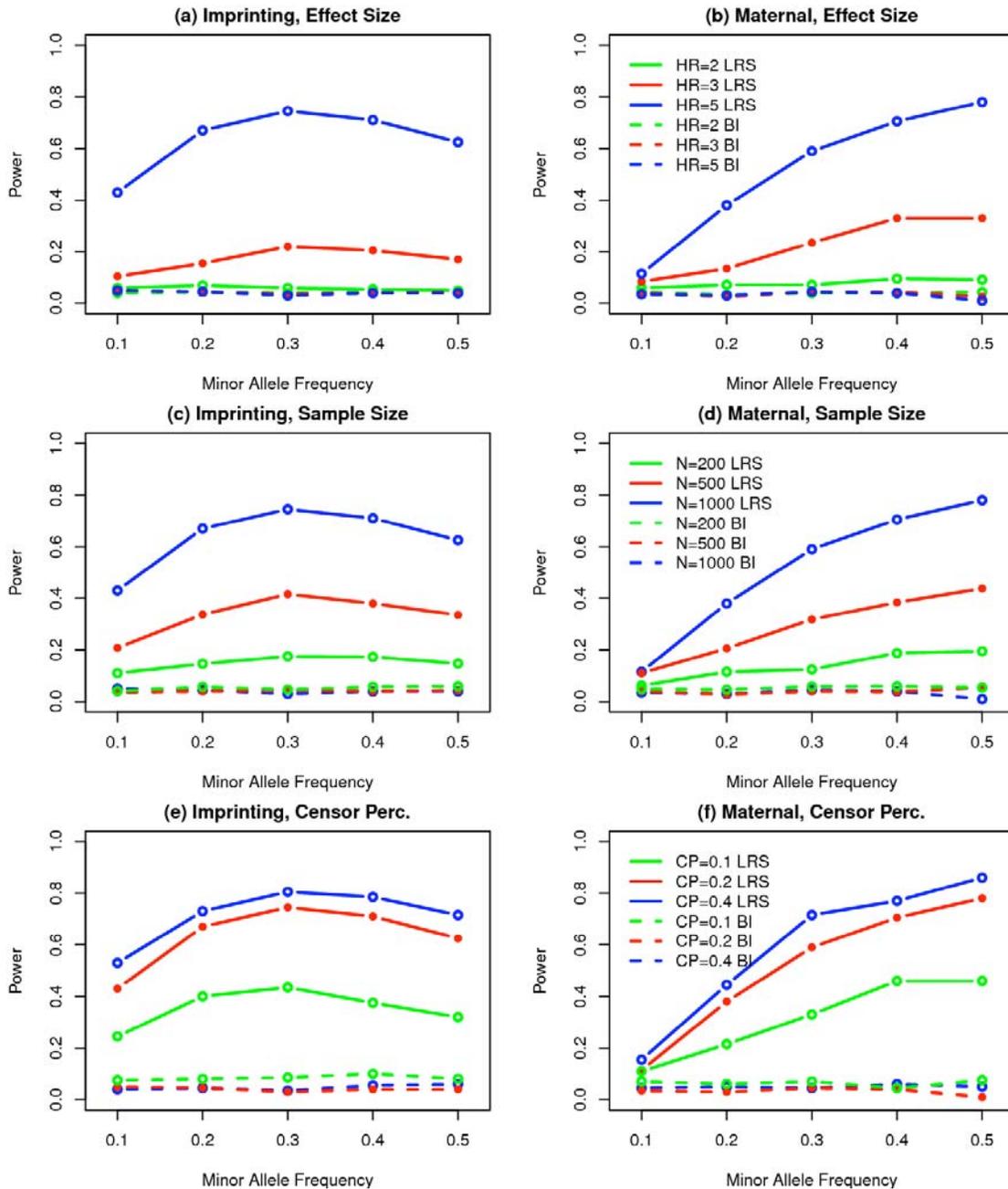


Figure 1: Power of detecting significant associations between the offspring's status and maternal history scores under different scenarios. Each left panel shares the same legend as the right panel. LRS is short for log-rank score and BI is for binary indicator.

We found that the test has limited power when the samples size is less than 1000 (Figure 1c, d) and the power almost reduces linearly with the proportion of censoring decreases (Figure 1e, f).

For single-locus models, our method has larger power to detect an imprinting effect than a maternal effect given the same effect size and MAF smaller than 0.4 and this reverses when the MAF of the casual allele is close to 0.5. In either scenario, the tests using LRS had remarkable power gain over those using the binary indicators. These suggest that we will require larger sample size to detect a single-locus maternal effect using our approach and it is extremely difficult for a binary indicator to claim a significance even for a large maternal effect with sample size around a few thousands.

Please note that a hazard ratio of 5 for 1 allele copy change corresponds to a much smaller hazard ratio (1.7-2.1) between two quartiles of the LRS or an even smaller hazard ratio (1.2-1.7) for a standard deviation change in the LRS. Therefore, the sample size required to have a decent power (>80%) for a small percentage

of censoring and small effect size is about 1100-2000 and would be higher with more censored observations when the true mechanism is single-locus imprinting [21].

Results from REGARDS

Among all proband, 10,747 (46.88%) were males and 12,176 (53.12%) were females; 58.27% were Caucasians and 41.73% were African Americans. The average age of the proband was 66.0 (s.d.=9.12) years. For their parents, the risk of having MI and stroke was age- and sex-dependent, increasing at younger ages and peaked at older ages, and then falling again at advanced age as shown in Figure 2.

Clearly, the incidence rates for both stroke and MI were different for fathers and mothers. Therefore, we estimated the proposed LRS separately for paternal and maternal history. Figure 3 shows the risk scores for stroke (left panel) and MI (right panel) at different ages for fathers (blue solid curve) and mothers (red dashed curves). The difference in stroke risk scores between mothers and fathers was small and almost a constant 0.05 between ages of onset 35-95. The MI risk scores,

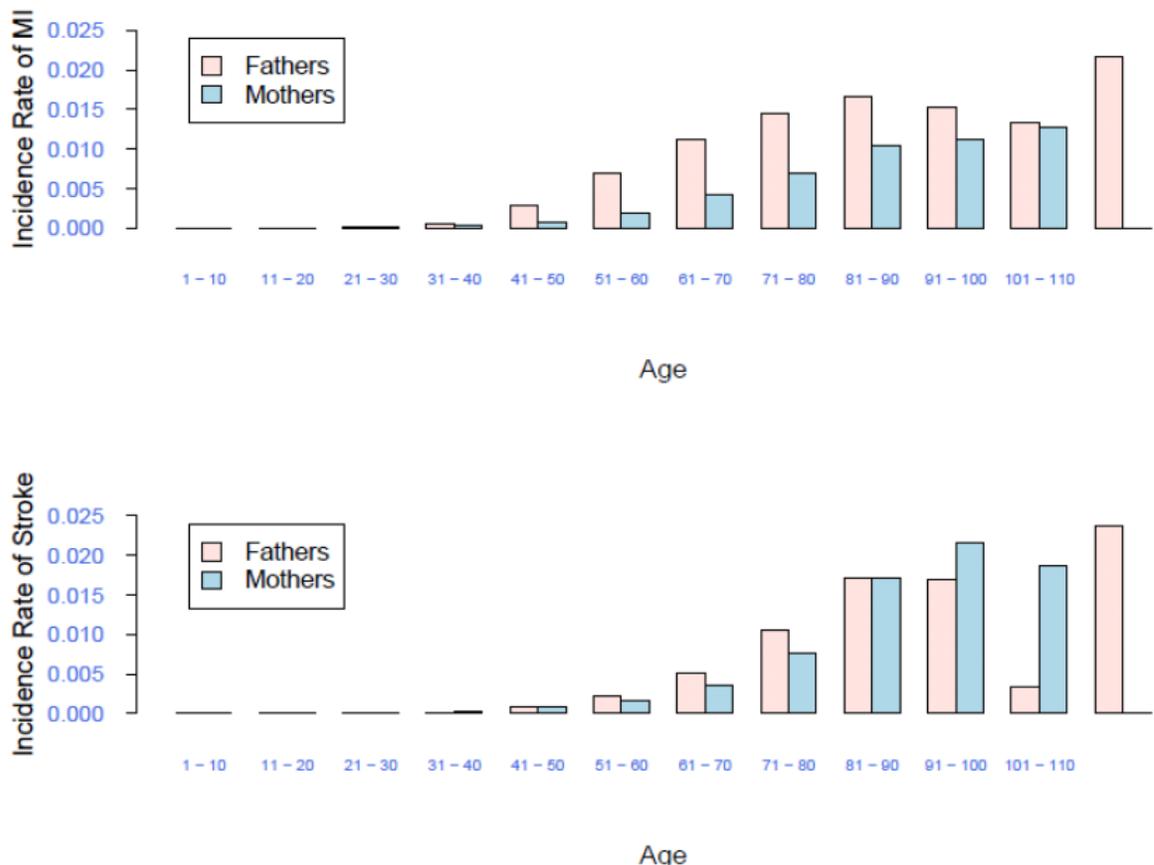


Figure 2: 10-year incidence rate (per person per year) of MI and stroke among fathers and mothers of the proband in REGARDS.

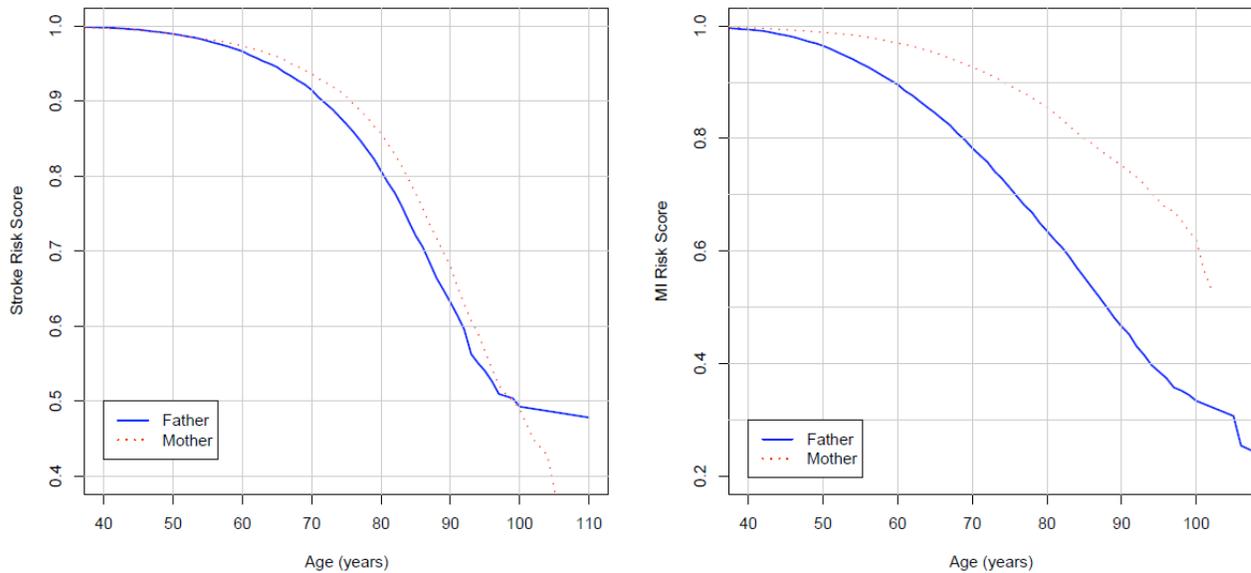


Figure 3: Stroke and MI risk curves for fathers and mothers.

however, differ significantly between mothers and fathers, with mothers having alarmingly higher risk scores at all ages, reflecting the lower MI incidence rate in females as shown in Figure 2. Since the REGARDS study is the largest cohort study on heart disease so far, we believe that the log-rank chart we have shown in Figure 3 could be used by doctors and researchers to assess stroke and MI risk for the general public.

For the proband, we calculated the survival curves of stroke for three groups separated by the upper and lower quartiles of their parental LRSs. The first group with LRS in the lower quartile has lowest risk from their parents who lived longer enough before having a stroke and the third group with LRS in the upper quartile corresponds to the highest risk whose parents had a stroke at much younger ages. Figure 4 shows all the calculated survival curves for different quartiles of the maternal and paternal LRSs. As a comparison, the survival probabilities of the offspring were also calculated separately for maternal and paternal stroke history using both the binary indicator and the LRS. Figure 4 suggests that the maternal history of stroke has a bigger influence than the paternal history on their offspring's survival. Especially when the patients were 60+ years old, whether the mother had a stroke at a younger age or an old age could affect up to 10% the chance of survival, or up to ten years of life expectancy at the same survival probability. This clearly demonstrates the advantages of using the log-rank score over the binary indicator. As shown at the bottom panels of Figure 4, the binary indicator could not give enough warning, especially at early age.

The same survival curves for MI are provided in Figure 5. Different from stroke, individual MI risks seem positively associated with both maternal and paternal MI histories. Also, the curves stratified by paternal history LRSs are further apart than using binary indicators, suggesting capability of desegregating risk groups.

The results from the final Cox models with significant covariates and parental health history for the proband are listed in Table 2.

For stroke, we have found a strong correlation with maternal stroke history ($p=4.7e-6$) yet a non-significant correlation with paternal stroke history ($p=0.29$) using LRS. Using the binary indicator, the p-values were 0.028 and 0.66, respectively. Both maternal and imprinting effects could give such results, as shown in our simulations. But, since the maternal effect generally influences early life disorders, we suspect that the imprinting effect was the main cause for the result.

The MI event of offspring is significantly associated with both parental disease histories, regardless of the offspring's gender.

The larger C-index for the models using our LRS vs. using binary indicators also confirm the merit of our proposed LRS though the difference is minor. For stroke, both models using maternal history had slightly larger C-index than those using paternal index.

To investigate whether there is a gender difference in the association pattern, we fit the same Cox model

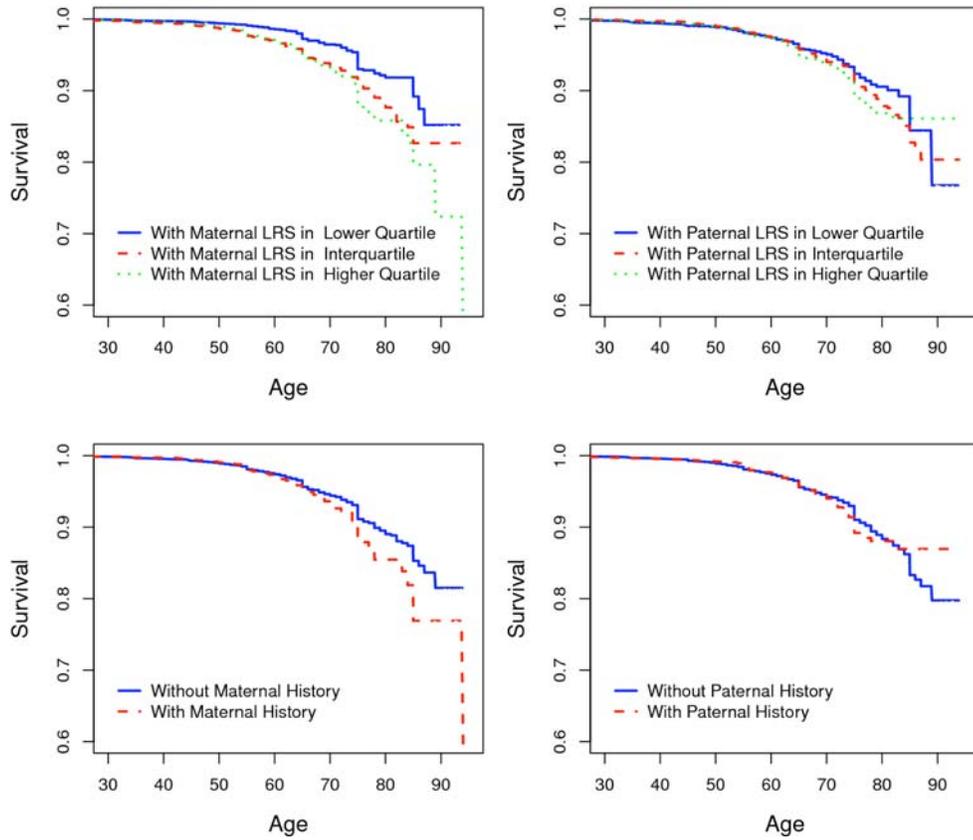


Figure 4: Stroke survival curves stratified by maternal and paternal LRSs and binary disease history indicators.

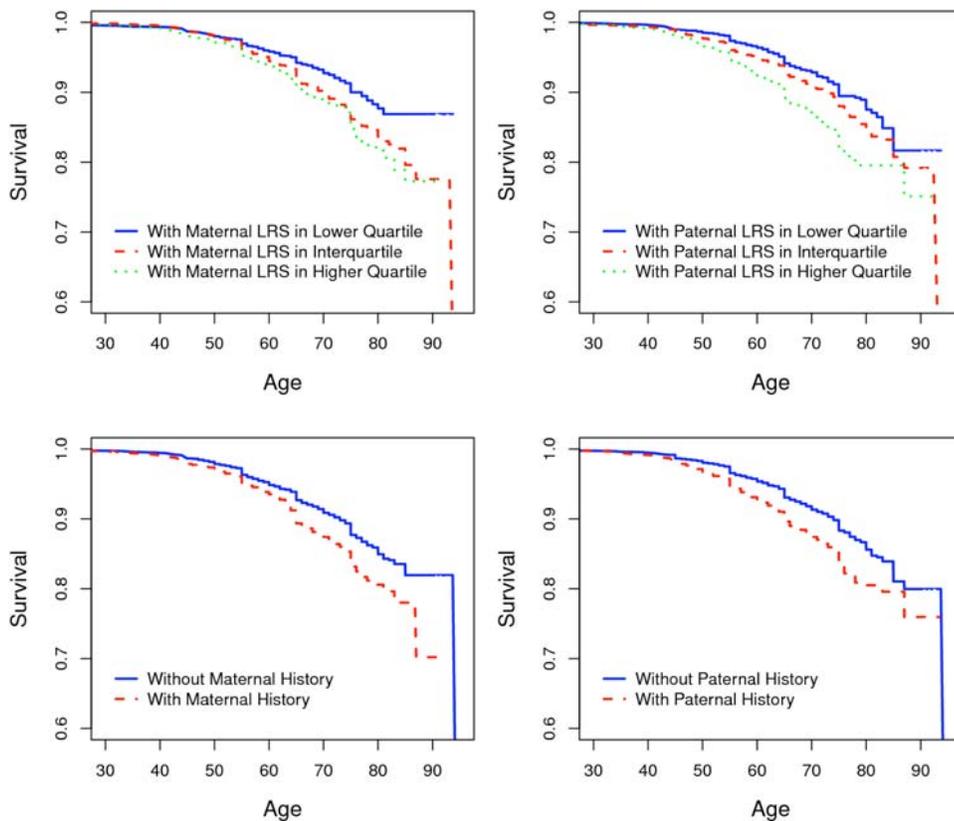


Figure 5: MI survival curves stratified by maternal and paternal LRSs and binary disease history indicators.

Table 2: Hazard Ratios (p-Values) in Cox PH Model Using Parental History and Other Risk Factors as Covariates

	In the model with			
	Maternal LRS	Paternal LRS	Maternal binary indicator	Paternal binary indicator
Event of Interest: Stroke				
Parental history [†]	1.346 (4.7e-6)	1.072(2.9e-1)	1.238 (2.8e-2)	1.048 (6.6e-1)
Age	0.922 (<1e-12)	0.923 (<1e-12)	0.920 (<1e-12)	0.923 (<1e-12)
Race (W)	0.785 (5.4e-3)	0.771 (2.8e-3)	0.770 (2.6e-3)	0.771 (2.8e-3)
Cholesterol level	0.998 (1.0e-1)	0.997 (4.9e-3)	0.998 (1.1e-1)	0.997 (4.9e-3)
HDL	0.988 (2.9e-5)	0.990 (3.0e-4)	0.988 (2.2e-5)	0.990 (2.9e-4)
Hypertension	1.968 (1.0e-12)	2.003 (3.7e-13)	1.995 (3.5e-13)	2.012 (2.4e-13)
Diabetes	1.687 (4.1e-9)	1.710 (1.3e-9)	1.704 (2.0e-9)	1.714 (1.1e-9)
C-Index	0.707	0.703	0.706	0.703
Log-likelihood	-5402.173	-5426.877	-5410.287	-5427.332
N	12486	12567	12486	12567
Event of Interest: Myocardial Infarction				
Parental history [†]	1.365 (1.7e-8)	1.344 (3.6e-9)	1.351 (3.2e-4)	1.347 (2.0e-5)
Age	0.974 (2.2e-7)	0.976 (6.2e-7)	0.973 (4.0e-8)	0.975 (1.8e-7)
Race (W)	1.455 (7.4e-6)	1.317 (9.7e-4)	1.424 (2.4e-5)	1.337 (5.0e-4)
Cholesterol level	0.990 (<1e-12)	0.990 (<1e-12)	0.990 (<1e-12)	0.990 (<1e-12)
HDL	0.979 (5.6e-16)	0.980 (2.3e-15)	0.979 (4.4e-16)	0.980 (2.1e-15)
ln(DBP)	0.585 (4.3e-2)	0.589 (4.5e-2)	0.575 (3.7e-2)	0.593 (4.8e-2)
ln(Crp)	1.074 (1.3e-2)	1.070 (1.8e-2)	1.077 (9.4e-3)	1.074 (1.2e-2)
Hypertension	1.440 (1.3e-6)	1.427 (2.0e-6)	1.458 (5.4e-7)	1.444 (8.8e-7)
Diabetes	1.555 (3.4e-9)	1.580 (5.6e-10)	1.563 (2.3e-9)	1.590 (3.3e-10)
Log-likelihood	-7858.71	-8002.214	-7868.062	-8010.577
C-Index	0.682	0.679	0.682	0.680
N	11006	11104	11006	11104

[†]: corresponding to the four maternal or paternal history scores respectively.

for male and female offspring separately and the results were similar to those using combined data, thus ruling out gender difference for both MI and stroke.

6. CONCLUSION AND DISCUSSION

In this study, we have proposed a log-rank score to assess parental histories of a disease. Unlike the binary indicator, which only accounts the parents' status of disease, the log-rank score takes parents' ages and relative risk into consideration. If a parent experienced disease at a relatively young age, his/her score would be high, and vice versa. This log-rank score can be considered as the binary indicator offset by factors related to age, and thus better reflects genetic risk.

We have applied the LRS separately to mothers and fathers in the REGARDS study and found that for stroke, the maternal LRSs are a little bit higher than the

paternal ones, but for MI, the maternal LRSs are substantially higher. Investigating the association between parental risk score and offspring's risk of disease, we have found that the risk of stroke is strongly associated with maternal history but not with paternal history, suggesting imbalance of parental effects, and the risk of MI is strongly associated with both maternal ($p=1.7e-8$) and paternal ($p=3.6e-9$) histories. Using the binary scores, we have found similar conclusions, albeit with much weaker signals.

We did sensitivity analysis, using a binary family history indicator defined as yes only when parental event occurs before age 60. The results were similar as using the simple binary indicator.

In cases where strong associations were found, the p-values obtained using the LRSs were 4 magnitudes lower than those obtained using the binary indicator; in cases of weak association, the p-values were similar

using either of them. Our simulation studies also suggest that the proposed LRSs are more likely to be found significantly associated with the offspring's disease than the binary indicators regardless of effect size and causal effect frequency, and that the LRSs yield conservative type I error rates comparable to those of binary indicators. Both the real application and the simulations suggest that the proposed LRS is well-behaved under the null hypothesis and is more powerful to detect associations with true underlying family risk across a broad range of alternative hypotheses.

It is known that genetic factors play an important role in heart diseases. Parental cardiovascular disease history has been suggested as an independent predictor of offspring cardiovascular events in middle-aged adults [22]. Moreover, it has been shown that a maternal history of myocardial infarction might be more strongly associated with risk of cardiovascular disease than paternal history [23]. Although our study gave both maternal (HR=1.365, $p=1.7e-8$) and paternal (HR=1.344, $p=3.6e-9$) history of MI basically the same weight of importance in general, the log-rank scores for mothers are indeed substantially higher than the fathers' at the same ages. This could possibly explain our observations.

Our analysis suggests that proposed LRS can better quantify the risks of parental history and help clinicians and patients with prevention of cardiovascular disease. This score only requires the parental disease and age-on-set information without burdening collection and analysis of genetic markers.

Because the disease status of family members were obtained based on children's recalls during interviews, the explicit forms of stroke for those affected family members were unknown. We cannot investigate the risk factor and heritability difference among stroke subtypes, which may limit the accuracy of risk prediction if the subtypes are heterogeneous.

In a study of cancer family history [24], the number of reported paternal family histories of cancer was much less than that of maternal cancer histories, which promotes careful inspection of reporting bias in similar studies. In our real study, both ratios of stroke and MI in mothers vs. that in fathers were the same as in the overall population [25] and there is no evidence of reporting bias.

For complex diseases that have a wide range of onset age and various causes, this study has

demonstrated the value of using the LRS to quantify parental disease history. In addition, LRS is a nonparametric score and requires minimum assumptions so it is flexible and relatively robust to various underlying distributions.

Alternatively, the Gehan-Wilcoxon score [26] can be used if earlier events need to be emphasized and later censoring are more likely to be caused by other factors not related with diseases. However, this score has lower power in the event occurring in later life than the log-rank score, as seen in our simulation (data not shown).

ACKNOWLEDGEMENT

The authors want to thank Dr. Lei Liu at Northwestern University for his suggestion on the simulation setup. R. Feng is supported by a research grant R01GM088566 from National Institute of Health and G. Howard is supported by a cooperative agreement U01 NS041588 from the National Institute of Neurological Disorders and Stroke, National Institute of Health.

REFERENCES

- [1] Bassett SS, Avramopoulos D, Fallin D. Evidence for parent of origin effect in late-onset Alzheimer disease. *Am J Med Genet* 2002; 114(6): 679-86.
<http://dx.doi.org/10.1002/ajmg.10648>
- [2] Bjornholt JV, Erikssen G, Liestol K, Jervell J, Thaulow E, Erikssen J. Type 2 diabetes and maternal family history: an impact beyond slow glucose removal rate and fasting hyperglycemia in low-risk individuals? Results from 22.5 years of follow-up of healthy nondiabetic men. *Diabetes Care* 2000; 23(9): 1255-9.
<http://dx.doi.org/10.2337/diacare.23.9.1255>
- [3] McCuaig JM, Greenwood CM, Shuman C, Chitayat D, Murphy KJ, Rosen B, Armel SR. Breast and ovarian cancer: the forgotten paternal contribution. *J Genet Couns* 20(5): 442-9.
- [4] Soroko SB, Barrettconnor E, Edelstein SL, Kritzsilverstein D. Family History of Osteoporosis and Bone-Mineral Density at the Axial Skeleton - the Rancho-Bernardo Study. *J Bone Mineral Res* 1994; 9(6): 761-69.
<http://dx.doi.org/10.1002/jbmr.5650090602>
- [5] Sun Y, Sherman M. Some permutation tests for survival data. *Biometrics* 1996; 52(1): 87-97.
<http://dx.doi.org/10.2307/2533147>
- [6] Howard G, Koch GG. The Glim Log-Rank Test - General Linear Modeling of Log-Rank Scores as a Method of Analysis for Survival-Data. *Commun Statistics-Simulation Comput* 1990; 19(3): 903-17.
<http://dx.doi.org/10.1080/03610919008812897>
- [7] Feng R, McClure LA, Tiwari HK, Howard G. A new estimate of family disease history providing improved prediction of disease risks. *Stat Med* 2009; 28(8): 1269-83.
<http://dx.doi.org/10.1002/sim.3526>
- [8] Bernstein IM, Plociennik K, Stahle S, Badger GJ, Secker-Walker R. Impact of maternal cigarette smoking on fetal

- growth and body composition. *Am J Obstet Gynecol* 2000; 183(4): 883-6.
<http://dx.doi.org/10.1067/mob.2000.109103>
- [9] Doege K, Grajecki D, Zyriax BC, Detinkina E, Zu Eulenburg C, Buhling KJ. Impact of maternal supplementation with probiotics during pregnancy on atopic eczema in childhood--a meta-analysis. *Br J Nutr* 107(1): 1-6.
- [10] De-Regil LM, Palacios C, Ansary A, Kulier R, Pena-Rosas JP. Vitamin D supplementation for women during pregnancy. *Cochrane Database Syst Rev* 2: CD008873.
- [11] van den Hooven EH, de Kluizenaar Y, Pierik FH, Hofman A, van Ratingen SW, Zandveld PY, *et al.* Chronic Air Pollution Exposure during Pregnancy and Maternal and Fetal C-reactive Protein Levels. The Generation R Study. *Environ Health Perspect*.
- [12] Shah R, Diaz SD, Arria A, Lagasse LL, Derauf C, Newman E, *et al.* Prenatal Methamphetamine Exposure and Short-Term Maternal and Infant Medical Outcomes. *Am J Perinatol*.
- [13] Singer LT, Moore DG, Fulton S, Goodwin J, Turner JJ, Min MO, Parrott AC. Neurobehavioral outcomes of infants exposed to MDMA (Ecstasy) and other recreational drugs during pregnancy. *Neurotoxicol Teratol*.
- [14] Honea RA, Swerdlow RH, Vidoni ED, Burns JM. Progressive regional atrophy in normal adults with a maternal history of Alzheimer disease. *Neurology* 76(9): 822-9.
- [15] Koch GG, Sen PK, Amara I. Log-rank scores, statistics, and tests. In *Encyclopedia of Statistical Sciences*, Kotz S, Johnson NL, Eds. John Wiley & Sons: New York, NY. 1985; pp. 136-142.
- [16] Koch GG, Sen PK, Amara I. Log-rank scores, statistics, and tests. In *Encyclopedia of Statistical Sciences*, Kotz S, Johnson NL, Eds. John Wiley & Sons: New York, NY. 1985; pp. 136-142.
- [17] Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA* 2007; 297(6): 611-9.
<http://dx.doi.org/10.1001/jama.297.6.611>
- [18] Harrell FE, Lee KL Jr., Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; 15(4): 361-87.
[http://dx.doi.org/10.1002/\(SICI\)1097-0258\(19960229\)15:4<361::AID-SIM168>3.0.CO;2-4](http://dx.doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4)
- [19] Gonen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 2005; 92(4): 965-70.
<http://dx.doi.org/10.1093/biomet/92.4.965>
- [20] Potapov S, Adler W, Schmid M. survAUC: Estimators of Prediction Accuracy for Time-to-Event Data 2011; Available from: <http://cran.at.r-project.org/web/packages/survAUC/index.html>
- [21] Hsieh FY, Lavori PW. Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Control Clin Trials* 2000; 21(6): 552-60.
[http://dx.doi.org/10.1016/S0197-2456\(00\)00104-5](http://dx.doi.org/10.1016/S0197-2456(00)00104-5)
- [22] Lloyd-Jones DM, Nam BH, D'Agostino RB, Levy D Sr., Murabito JM, Wang TJ, *et al.* Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* 2004; 291(18): 2204-11.
<http://dx.doi.org/10.1001/jama.291.18.2204>
- [23] Sesso HD, Lee IM, Gaziano JM, Rexrode KM, Glynn RJ, Buring JE. Maternal and paternal history of myocardial infarction and risk of cardiovascular disease in men and women. *Circulation* 2001; 104(4): 393-8.
<http://dx.doi.org/10.1161/hc2901.093115>
- [24] Ozanne EM, O'Connell A, Bouzan C, Bosinoff P, Rourke T, Dowd D, Drohan B, *et al.* Bias in the reporting of family history: implications for clinical care. *J Genet Couns* 21(4): 547-56.
- [25] Pal S. Prevalence of Selected Cardiovascular Diseases US Pharm 2009; 34(2): 2009.
- [26] Gehan EA. A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. *Biometrika* 1965; 52(1/2): 203-23.

Received on 26-06-2013

Accepted on 14-11-2014

Published on 31-01-2014

<http://dx.doi.org/10.6000/1929-6029.2014.03.01.4>© 2014 Feng *et al.*; Licensee Lifescience Global.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.