Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/040202

Comparisons between Survival Models in Predicting Cardiovascular Disease Events: Application in the ATTICA Study (2002-2012).

Ekavi N. Georgousopoulou[1], *Christos Pitsavos*[2], *Christos Mary Yannakoulia*[1] *and Demosthenes B. Panagiotakos* [1,*].

[1] Department of Nutrition and Dietetics, School of Health Science & Education, Harokopio University, Athens, Greece.
[2] First Cardiology Clinic, School of Medicine, University of Athens, Greece.

Received: 26 Feb. 2015, Revised: 7 Jun. 2015, Accepted: 8 Jun. 2015
Published online: 1 Jul. 2015

Abstract: In order to assess individual's risk of a disease, semi-parametric proportional hazards Cox models are mostly used, while fewer studies have used parametric models. The aim of the present work was to compare semi-parametric and parametric statistical methods regarding their goodness of fit. To investigate the research hypothesis, characteristics of the 3042 participants of the ATTICA epidemiological study, were used; 2583 of them were found in the 10-year follow-up (2011-2012) and 317 (15.7%) developed a cardiovascular disease event. Three multivariable models, adjusted for the same set of risk factors were compared regarding their performance, using the Bayesian Information Criterion (BIC). All models were adjusted for: age, gender, physical activity level, Body Mass Index, smoking, hypertension, diabetes mellitus, hypercholesterolemia and adherence to the Mediterranean dietary pattern (assessed with MedDietScore). The semi-parametric Cox proportional hazard model had the worst performance as compared with the parametric survival models under the Weibull distribution. Between the two other parametric models, the Weibull model had the best performance (BIC =1386.488) as compared with the model with the exponential distribution (BIC =1729.724) (p for Harell?s C ¡0.001). It appears that parametric models in relation to semi-parametric Cox proportional hazard models have better performance, while parametric model with Weibull distribution had the best performance among the parametric models.

Keywords: Survival analysis, parametric models, semi-parametric models, model performance, cardiovascular risk.

1 Introduction

The latest guidelines for cardiovascular diseases (CVD) prevention, strongly suggested the estimation of individual 10-year CVD risk for all adult subjects, independently of their health profile, in order to better treat the disease. CVD risk estimation scores have not only widely applied to everyday clinical practice, but suggests an important component of evidence-based medicine [1]. The CVD risk estimation scores provide an individual CVD risk estimation for a specific future time period ?usually decade as regards to CVD-, using easily assessed parameters such as, gender, age, smoking status, history of diabetes mellitus, hypertension, hypercholesterolemia [2]. The accuracy of these scores has raised several discussions, underlying a challenge for modern CVD epidemiology, i.e., the optimization of CVD risk estimation scores, in order to better identify subjects at high CVD risk, without simultaneously treating more healthy people [3]. In the existing CVD risk prediction models various statistical techniques have been used, mainly due to the methodological differences of the existing studies. E.g. some models have used risk factors? scores and give information only regarding the absolute individual risk, whereas other scores provide estimation of relevant CVD risk scores, as compared with subjects without known CVD risk factors. The statistical approaches vary between CVD risk estimation models, which use parametric models (i.e., Weibull) that requires the evaluation of baseline hazard, in contrast with semi-parametric Cox proportional hazards models, which do not require the assessment of baseline hazard [4,5]. Despite the fact that several methodological approaches have been used, it remains unknown whether any of them performs better in estimating CVD risk. The increasing interest for CVD risk estimation scores in everyday clinical practice, lead to an emerging need of upgraded estimation scores that would correctly classify subjects at high CVD risk [6]. Therefore, the present study aimed

* Corresponding author e-mail: d.b.panagiotakos@usa.net

to evaluate which of the existing statistical methodologies for survival data (Cox proportional hazard models, exponential distribution or Weibull distribution) performs better to the CVD risk estimation.

2 Methods

2.1 Statistical Methodology

The semi-parametric proportional hazards Cox model [7] (see (1), h_i =hazard rate) suggests that the hazards ratio remains constant over time:

$$\log h_i(t) = \log h_0(t) + \beta_1 x_{i1} + ... + \beta_k x_{ik} \quad (1)$$

The Cox proportional hazards model does not estimate the baseline hazard. Contrarily, the exponential distribution with λ parameter, assumes that the hazards ratio remains constant over time, but, the estimation of baseline hazard is possible through the distribution parameters, given from 2, where $\mu = log(\lambda)$ [8].

$$\log h_i(t) = \mu + \beta_1 x_{i1} + ... + \beta_k x_{ik} \quad (2)$$

As regards the Weibull distribution (with λ and k parameters), the proportionality of hazard is assumed, but the hazards ratio can be estimated by 3, and depend on time in various ways [9]:

$$\log h_i(t) = \mu + \alpha \log t + \beta_1 x_{i1} + ... + \beta_k x_{ik} \quad (3)$$

Estimations are performed using the Maximum Likelihood methodology.

2.2 Sampling procedure at baseline examination

The working sample used to test the research hypothesis was the ATTICA study. In brief, the study was carried out in the greater metropolitan Athens area (including 78% urban and 22% rural regions) during 2001-2002. Of the 4056 invited individuals, 3042 agreed to participate (75% participation rate); 1514 of the participants were men (18-87 years) and 1528 were women (18-89 years). Exclusion of CVD at baseline examination was performed through a detailed clinical evaluation by the physicians of the study that followed standard criteria. The examination was performed in the individuals? homes or workplaces places. The baseline evaluation of the ATTICA study included information about: socio-demographic characteristics, history of hypertension, hypercholesterolemia and diabetes, family history of CVD, dietary and other lifestyle habits (i.e., smoking status and physical activity). The recorded variables were gender, age, Body Mass Index (BMI) (using the measured weight and height and the formulae for BMI proposed by Lambert Adolphe Jacques Quetelet (1796-1874), i.e., body weight (in kilograms/ height (in meters)2), the physical activity status was assessed using the Greek validated version of the International Physical Activity Questionnaire (IPAQ); participants were classified as physically active or sedentary [10]. Arterial blood pressure was measured at the end of the baseline physical examination with subject in sitting position, and at least 30 minutes at rest. Participants whose average blood pressure levels were greater or equal to 140 / 90 mmHg or were under antihypertensive medication were classified as having hypertension. Hypercholesterolemia was defined as total cholesterol levels greater than 200 mg/dl or the use of lipids lowering agents. Blood glucose levels (in mg/dL) were measured with a Beckman Glucose Analyzer (Beckman Instruments, Fullerton, CA, USA). Diabetes mellitus (type 2) was defined according to the American Diabetes Association diagnostic criteria (i.e., blood glucose levels greater than 125 mg/dL classified participants as having diabetes). Smokers were defined as those who were smoking at least one cigarette per day during the past year or had recently stopped smoking (during a year); the rest of the participants were defined as non-smokers. The MedDietScore was used to evaluate adherence to the Mediterranean diet. More specifically, MedDietScore is an index with 11 questions regarding consumption frequency of the main foods of the Mediterranean diet pyramid. In particular, based on the suggested intake, monotonic functions (with the exception of alcohol intake) were used in order to score the frequency consumption of these foods. In particular, individual ratings (from 0 to 5 or the reverse) were assigned in each of the 11 food groups according to their position in the Mediterranean diet pyramid. Thus, the score ranges from 0 to 55, with higher scores indicating greater level of adherence to the Mediterranean diet [11]. Further details about the working sample used here may be found elsewhere [12].

2.3 Follow-up examination (2011-2012)

During 2011-12, the ATTICA Study?s investigators performed the 10-year follow-up (mean follow-up time 8.41 y). Of the n=3042 initially enrolled participants, n=2583 were found during the follow-up (85% participation rate). Of the individuals that were lost to follow-up (i.e., n=459), n=224 were not found because of missing or wrong addresses and telephone numbers that they have provided at baseline examination and n=235 because they denied being re-examined. No differences were reported regarding the distribution of sex (men 50% vs. 49%, p=0.613), obesity (19% vs. 16%, p=0.208), as well as anxiety (p=0.083) and depression levels (p=0.173) between the participants that were found to follow-up and the participants that were lost to follow-up. All cases with missing information were excluded from the analyses. Thus, for the present work, complete data from n=2009 participants with CVD evaluation at follow-up were used. In order to participate in the follow-up all participants were initially appointed through telephone calls. Afterwards, the investigators approached the participants and performed a detailed evaluation of their medical records. Among others, information about participants': (a) vital status (death from any cause or due to CVD), (b) development of CHD (i.e., myocardial infarction, angina pectoris, other identified forms of ischemia -WHO-ICD coding 410-414.9, 427.2, 427.6-, heart failure of different types, and chronic arrhythmias -WHO-ICD coding 400.0-404.9, 427.0 -427.5, 427.9-), and (c) development of stroke (WHO-ICD coding 430-438), was assessed and considered as the outcome in this work.

2.4 Statistical analysis

Continuous variables were presented as mean valuesstandard deviation and categorical variables are presented as frequencies. The continuous variables were tested for following the Normal distribution through P-P plots. Associations between categorical variables were tested using the chi-square test. Comparisons between mean values of normally distributed variables between those who developed an event and the rest of the participants were performed using Student?s t-test, after controlling for equality of variances using the Levene?s test. The time to CVD event was recorded on annual basis. Log-rank test was also applied to evaluate differences between groups of participants as regards CVD incidence. The hazard ratios (HR) of developing a CVD event during the 10-year period, according to the participants? baseline characteristics (i.e., age, adherence to Mediterranean diet (MedDietScore), physical activity status, smoking, history of hypertension, diabetes and hypercholesterolemia) were estimated using (a) Cox proportional hazards models, (b) Exponential and (c) Weibull survival models. For testing the assumption of proportional hazards, graphical control was applied between the classes of the categorical variables (gender, hypertension, hypercholesterolemia and diabetes), using the command stphtest in STATA version 13.0. Moreover, the log-log plots were used to test the assumptions for both the Exponential and the Weibull models. For all models the Bayesian information criterion (BIC) was also calculated [13]:

$$BIC = -2 * lnL + k * ln(n) \tag{4}$$

(L the likelihood of the models, k the degrees of freedom and n the study sample), which suggests a standard criterion for the good performance of the model to the observed data. As the BIC decreases, the model performs better to observed data. For the comparison among parametric and semi-parametric (Cox) models, the graphical control of Cox & Snell residuals in relation to cumulative hazard was applied[14] Specifically, the Cox & Snell residuals of each model were calculated through formulae (5):

$$rcs_i = exp(?x_i)?_0(t_i) \tag{5}$$

Then, the residuals were presented graphically with Nelson & Aalen cumulative hazard [15]. The closest the residual to the diagonal line of $45^?$ that starts form (0,0), the best performs the model to the observed events. Under the assumption of proportional hazards the cumulative distribution should follow the diagonal.

Under the assumption of proportional hazards the cumulative distribution should follow the diagonal. Harrell?s C and the equivalent parametric Somers? D have been proposed as measures of the predictive power of a general regression model and used in order to compare the discriminant ability of the models [15]. Harrell?s C is defined as the proportion of all pairs in which the predictions and outcomes are concordant [16]. The aforementioned estimator is a rank parameters? estimator, in the family of Kendall-? coefficient. As regards Somers? D, paired data (X_i, Y_i) and data sampling $(X_i, Y_i), (X_j, Y_j)$ from the relevant sampling population, have two possible outcomes: concordant and discordant pairs. A couple of observed data is concordant when higher X values is associated with higher Y values, but discordant when higher X values are associated with lower Y values. Thus, Somers? D(X — Y) estimator is the difference between conditional probabilities and Harrell?s C(X — Y) estimator is D(X — Y)+1/2. Both Harrell's C and Somers' D play an important role in rank statistics, and have been extensively used for assessing prediction performance in survival analysis settings. In the specific analysis, ?, ? represent the estimated and the observed CVD risk, respectively. The 95%

confidence intervals (CI) for the difference between Harrell?s C estimators were used instead of the 95% CIs of the estimators, due to the fact that the latter might become skewed in case of strongly positive associations [16]. The Harrell?s C estimations for all multivariable survival models were performed using STATA 11.0 software (StataCorp College Station, Texas 77845 USA). All other statistical analyses were performed using SPSS 18.0 software (SPSS Inc., Chicago, Il, USA).The statistical significance level was defined at a=0.05.

3 Results

3.1 10-year cardiovascular disease incidence

The, fatal or non-fatal, 10-year CVD incidence rate was n=317 (15.7%); of them, n=198 (19.7%) cases were men and n=119 (11.7%) cases were women. Of the n=317 CVD events, 46 were fatal (n=34 men), and, thus, the overall 10-year fatal CVD rate was 1.8% (3.4% for men and 1.2% for women). Based on the observed person-years, the annual incidence of CVD was 182 new cases per 10,000 men and 110 new cases per 10,000 women participants. The CVD mortality rate among men was almost 3 times greater than the same rate among women (3.34% vs. 1.2% respectively). As regards the non-fatal CVD events, men had an almost 2-fold greater 10-year incidence than women (16.19% vs. 9.83% respectively).

Table 1: Characteristics of the ATTICA study?s participants (n=2009) according to the 10-year fatal or non-fatal incidence of CVD.

		Status at 10-year follow up		
	Baseline	**CVD event free**	**CVD events**	**p**
		(n=1702)	(n=317)	
Age, yrs	4514	4313	5813	<0.001
Male gender, %	50%	48%	63%	<0.001
Smoking at,baseline or before, %	43%	55%	57%	0.462
Physical,activity, %	41%	41%	41%	0.999
MedDietScore,(0-55)	267	266	237	<0.001
Hypertension, %	30%	28%	51%	<0.001
Hypercholesterolemia,,%	39%	40%	57%	<0.001
Diabetes, %	7%	5%	22%	<0.001
Body mass index,,kg/m2	265	265	285	<0.001

P-values derived using the chi-square test or t-test for the categorical and continuous variables, respectively.

Characteristics of the participants by CVD status at 10-year are presented in Table 1. As it can be seen, the group of participants who developed CVD consisted by older, men, with increased blood lipids, systolic/diastolic blood pressure, glucose and C-reactive protein levels, as well as with lower adherence to the Mediterranean diet (i.e., lower diet score) and increased body mass index (all p-values ¡ 0.001). The three multi-adjusted survival models (i.e., semi-parametric Cox proportional hazards model, Weibull parametric model and exponential parametric model) for 10-year CVD risk are presented in Table 2. Specifically, the characteristics that were positively associated with higher risk of developing CVD within a decade were increasing age (Hazard Ratio (HR) for 1 year =1.05 (95%CI 1.04-1.07) for Cox model, 1.06 (1.05-1.07) for Weibull model and 1.05 (1.04-1.06) for exponential model), the male gender (HR male vs. female=1.56 (1.21-2.01) for Cox model, 1.57 (1.22-2.01) for Weibull model and 1.50 (1.17-1.93) for exponential model), Body Mass Index (HR for 1 kg/m2=1.02 (1.00-1.05), 1.02 (0.99-1.05) for Weibull model and 1.03 (1.00-1.06) for exponential model), hypercholesterolemia (HR for history vs. no history= 1.43 (1.13-1.80) for Cox model, 1.45 (1.15-1.84) for Weibull model and1.40 (1.11-1.77) for exponential model) and diabetes (HR for history vs. no history=1.73 (1.13-1.80) for Cox model, 1.73 (1.29-2.31) for Weibull model and 1.66 (1.24-2.22) for exponential model).

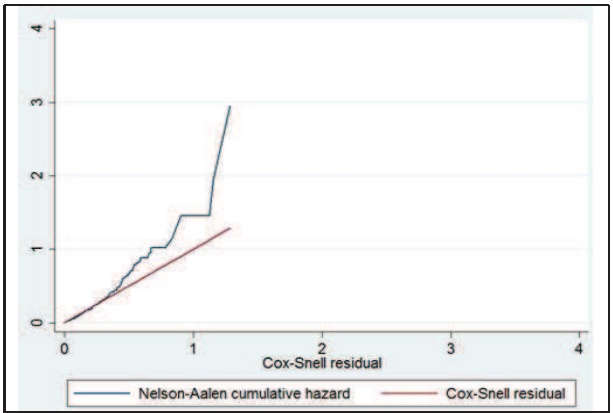
The two parametric models (i.e., Weibull and exponential) were further compared as regards to their performance to the observed CVD events. The model assuming the Weibull distribution had better performance than the model assuming exponential distribution (BIC= 1386.488 vs. BIC=1729.724 respectively, p for Harell?s C¡0.001). Moreover, the best parametric model (i.e., Weibull) was compared graphically with Cox proportional hazards model, using Cox-Snell residuals. The multivariable model assuming the Weibull distribution had better performance than the semi-parametric Cox model (Graph 1).

Table 2: Hazard ratios (95% Confidence Intervals) for multivariable parametric and semi-parametric survival models for development of cardiovascular disease, among the subjects of the ATTICA study (n=2009).

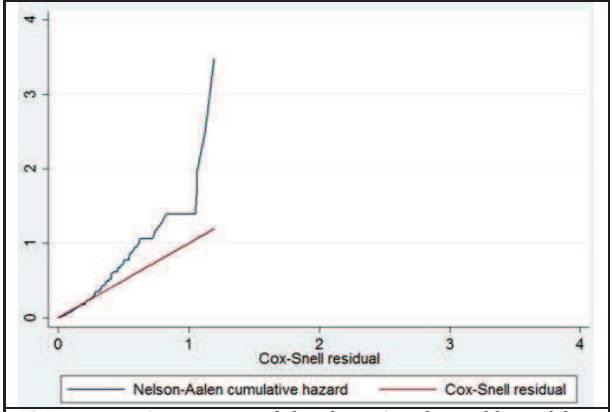
Variable	Cox	Weibull	Exponential
Age,(per 1 year)	1.05 (1.04-1.07)	1.06 (1.05-1.07)	1.05 (1.04-1.06)
Gender,(male vs. female)	1.56 (1.21-2.01)	1.57 (1.22-2.01)	1.50 (1.17-1.93)
Body Mass Index,(per 1 kg/m2)	1.02 (1.00-1.05)	1.02 (0.99-1.05)	1.03 (1.00-1.06)
Physical,activity (yes vs. no)	1.01 (0.80-1.29)	1.01 (0.79-1.28)	1.02 (0.80-1.29)
Smoking (yes vs.,no)	1.22 (0.95-1.56)	1.22 (0.95-1.56)	1.22 (0.95-1.56)
Hypertension,(yes vs. no)	1.21 (0.95-1.53)	1.21 (0.96-1.54)	1.22 (0.96-1.55)
Hypercholesterolemia,(yes vs. no)	1.43 (1.13-1.80)	1.45 (1.15-1.84)	1.40 (1.11-1.77)
Diabetes,(yes vs.,no)	1.73 (1.13-1.80)	1.73 (1.29-2.31)	1.66 (1.24-2.22)
MedDietScore,(per 1/55 unit)	0.99 (0.97-1.01)	0.99 (0.97-1.01)	0.99 (0.97-1.01)

4 Discussion

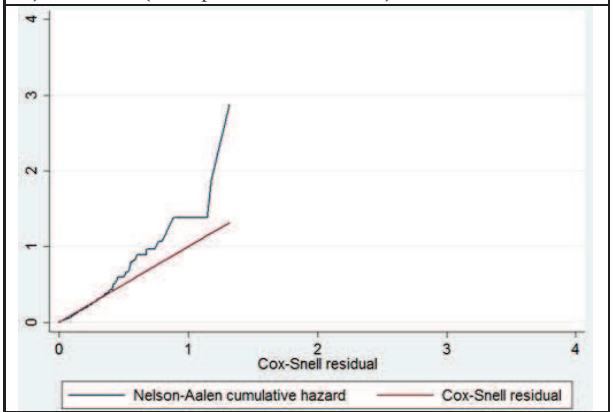
The present work aimed to compare the model performance of three different statistical methodologies that have been suggested for use in CVD risk prediction models and scores and to provide further insights in accurately predicting the risk for chronic diseases. According to the presented results, the parametric model under the assumption of the Weibull distribution had better performance to the observed CVD events than the semi-parametric Cox proportional hazards model and the parametric survival model assuming the Exponential distribution. Moreover, in terms of model performance, none of the aforementioned models had satisfactory prediction ability, thus, more research is essential in order to reveal the most important risk factors and the most appropriate methodology for estimating CVD risk. The aforementioned finding could be mainly attributed to the fact that Weibull parametric survival model has a crucial difference when compared with the other two survival models. Taking into account that increasing age consist a major CVD risk factor the assumption of the Weibull distribution, which empowers the aggravating effect of age on CVD risk, seems to have a better theoretical performance on CVD patho-biology. Specifically, the Weibull distribution assumes that times affect the risk for the outcome with a non-linear way. This finding bears physio-biological explanation for CVD risk, as increasing age was one the first established aggravating factors for the development of CVD. Relevant attempts have been performed by Cox et al. that compared the (Cox regression and generalized gamma survival models in two groups of patients (1504 males and 461 females) after their clinical diagnosis of AIDS. It was proved that generalized gamma model was better than Cox model because the assumption of proportional hazards was not fulfilled [17]. Moreover, Ravangard et al., compared the performance of parametric and semi-parametric models on the length of stay in hospital for subjects who recovered. The researchers concluded that as regards length of stay in hospital, the assumption of proportionality of hazards was not fulfilled and thus, proposed other parametric approaches [18]. The parametric and non-parametric approaches in survival analysis have important differences and their performances in observed data have raised researchers? interest lately [19]. The use of parametric models allows the projection of estimations in time-periods greater than the study?s follow-up period, which could be an important tool in predicting future risk for everyday clinical practice. Another advantage in using parametric survival models is that the baseline hazard can be directly estimated, whereas in Cox models this should be post estimated. Thus, the results of particular risk model can be directly applied to other datasets, whereas this is not possible using the Cox model unless it is supplied with the baseline hazard function. Nevertheless, using absolute risk estimations subjects could receive information about their actual CVD risk, not the relative CVD risk [20]. This information could be even more initiating for subjects at high risk in order, for example, to adopt a healthier lifestyle or comply with the proposed therapy. However, it should be underlined that the baseline hazard from another study may not be accurate to apply elsewhere, since it is biased by the sample used. However, for many populations where no prospective data are available, the use of parametric models could be an important alternative choice. Moreover, it should be mentioned that since the Weibull is a particular case of the Cox model and the Exponential is a case of the Weibull model, any discrepancies between the Weibull and Exponential suggest that the exponential model is not valid. Similarly discrepancies between the Weibull and the Cox suggest the Weibull is not valid. Based on the results presented in Table 2, it could be suggested that the model anyone may chose, has little impact on the estimated hazard ratios. The present study has several strengths, as it was attempted to methodologically approach the issue of CVD risk prediction, but, also some limitations, too. Baseline evaluation was performed once, which suggest that measurement error could bias the results. Thus, the prevalence of all clinical characteristics may have been overestimated. In addition the use of time-to-event in discrete values may also have limited the performance of the estimated models. Only three proportional hazards models were investigated here. Although, these three models are the most frequently used in CVD risk estimation and the discussions about their accuracy in prediction could provide important information for cardiovascular research, as well as everyday clinical



I) Semi-parametric multivariable Cox proportional hazards model



II) Parametric,(i.e. exponential distribution) multivariable model



III) Parametric,(i.e. Weibull distribution) multivariable model

Graph 1. Cox & Snell residuals for the 10-year incidence of CVD, for the three models' used.

practice, other parametric models (like the Gompertz model, lognormal model, etc) should also be used and compared. Moreover, taking into account that the exponential and the Weibull are particular cases of the Cox model, and if the Cox model is not valid, then neither are the other two, this suggests another limitation when aiming to compare model's performance. Additionally, the Exponential model may be too simplistic when age is considered, but it should be mentioned that all the three models included age as a covariate. As regards age as time, the Weibull model assumes a particular form for the association, whereas the Cox model does not. Thus, if the Weibull model fit the data well, may be more appropriate and preferred over the Cox model, since it makes better use of the data and is fully parametric, but this does not necessarily indicate that the Cox model is inadequate.

5 Conclusion

The use of parametric Weibull survival models seems to be a better approach than using semi-parametric Cox proportional hazards models in predicting future CVD events. Although, more research is needed to confirm or refute this finding, as well as bootstrapping methods for reducing bias, the aforementioned results may provide a useful mean for better identifying the potential CVD candidate in various populations.

Acknowledgements

The authors would like to thank the Reviewers of the Journal for the criticism made that helped to improve the presentation of the work done. Moreover, the authors would like to thank the ATTICA study group of investigators: Yannis Skoumas, Natassa Katinioti, Labros Papadimitriou, Constantina Masoura, Spiros Vellas, Yannis Lentzas, Manolis Kambaxis, Konstadina Palliou, Vassiliki Metaxa, Agathi Ntzouvani, Dimitris Mpougatsas, Nikolaos Skourlis, Christina Papanikolaou, Georgia-Maria Kouli, Aimilia Christou, Adella Zana, Maria Ntertimani, Aikaterini Kalogeropoulou, Evangelia Pitaraki, Alexandros Laskaris, Mihail Hatzigeorgiou and Athanasios Grekas for their assistance in the initial physical examination and follow-up evaluation, Efi Tsetsekou for her assistance in psychological evaluation, as well as the laboratory team: Carmen Vassiliadou and George Dedoussis (genetic analysis), Marina Toutouza-Giotsa, Constadina Tselika and Sia Poulopoulou (biochemical analysis) and Maria Toutouza for the database management.

Conflict of Interest

None to declare

Source of Funding

The ATTICA study was supported by research grants from the Hellenic Cardiology Society (HCS2002) and the Hellenic Atherosclerosis Society (HAS2003). Demosthenes Panagiotakos and Ekavi Georgousopoulou have received research grants by the Coca-Cola Company (2013).

References

[1] Perk J, G De Backer, H Gohlke, I Graham, Z Reiner, M.W.M. Verschuren, C Albus, P Benlian, G Boysen, R Cifkova, C Deaton, S Ebrahim, M Fisher, G Germano, R Hobbs, A Hoes, S Karadeniz, A Mezzani, E Prescott, L Ryden, M Scherer, M Syvanne, W.J.M. Scholte Op Reimer, C Vrints, D Wood, J-L Zamorano, F Zannad. European guidelines on cardiovascular disease prevention in clinical practice (version 2012). *Eur Heart J.* **33**: 1635-701, (2012).
[2] Panagiotakos D. Health measurement scales: methodological issues. *Open Cardiovasc Med J.* **3**:160-5, (2009).
[3] Cooney MT, Dudina AL, Graham IM. Value and limitations of existing scores for the assessment of cardiovascular risk: a review for clinicians. J Am Coll Cardiol. **54**, 1209-27, (2009).
[4] Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation, **97**, 1837-1847, (1998).
[5] Conroy RM, Pyrl K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetire P, Jousilahti P, Keil U, Njlstad I, Oganov RG, Thomsen T, Tunstall Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmsen L, Graham IM; SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J.* 24:987-1003, (2003).
[6] Graham IM. The importance of total cardiovascular risk assessment in clinical practice. *Eur J Gen Pract.*, **12**, 148-55, (2006).
[7] Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B.* **34**: 187?220, (1972).
[8] Devroye L. Non-Uniform Random Variate Generation. New York: Springer-Verlag. (1986).
[9] Weibull W. A statistical distribution function of wide applicability. *J. Appl. Mech.-Trans.* **18**:293?7, (1951).
[10] International Physical Activity Questionnaire (accessed at http://www.ipaq.ki.se/ on February 28, 2006).
[11] Panagiotakos DB, Pitsavos C, Stefanadis C. Dietary patterns: a Mediterranean diet score and its relation to clinical and biological markers of cardiovascular disease risk. Nutr Metab Cardiovasc Dis. **16**, 559-68, (2006).
[12] Pitsavos C, Panagiotakos DB, Chrysohoou C, Stefanadis C. Epidemiology of Cardiovascular risk factors in Greece; aims, design and baseline characteristics of the ATTICA study. *BMC Public Health.* **3**, 32:1-9, (2003).
[13] Schwarz GE. Estimating the dimension of a model. Annals of Statistics, **6**:461?4, (1978).
[14] Royston P. Explained variation for survival models. *The Stata Journal*, **6**, 83-96, (2006).

[15] Newson BR. Comparing the predictive powers of survival models using Harrell's C or Somers' D. *The Stata Journal*, **10**, 339-358, (2010).
[16] Daniels HE, Kendall, MG. The Significance of Rank Correlation Where Parental Correlation Exists. *Biometrika*, 34:197-208, (1947).
[17] Cox C, Chu H, Schneider MF, Muoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat Med*, **26**, 4352-74, (2007).
[18] Ravangard R, Arab M, Rashidian A, Akbarisari A, Zare A, Zeraati H. Comparison of the results of Cox proportional hazards model and parametric models in the study of length of stay in a Tertiary Teaching Hospital in Tehran, Iran. *Acta Medica Iranica*, **49**, 650-58, (2011).
[19] Wang W, Small D. A comparative study of parametric and nonparametric estimates of the attributable fraction for a semi-continuous exposure. *Int J Biostat*, **8**, 32, (2012).
[20] Rausand, M.,Hoyland, A. System Reliability Theory: Models, Statistical Methods, and Applications. 2004; Hoboken: John Wiley & Sons.