Estimation and development of 10- and 20-year cardiovascular mortality risk models in an industrial male workers database

Gil Harari *, Manfred S. Green, Shira Zelber-Sagi

School of Public Health, Faculty of Social Welfare and Health Sciences, University of Haifa, Haifa, Israel

A R T I C L E I N F O

Article history:
Received 12 February 2017
Received in revised form 4 June 2017
Accepted 12 July 2017
Available online 18 July 2017

Keywords:
Coronary heart disease
Cardiovascular disease
Atherosclerotic cardiovascular disease
Risk scores
Prediction models
Industrial cohort

A B S T R A C T

We examined the performance of the Framingham Heart Study (FHS) and the European Systematic Coronary Risk Evaluation (SCORE) models for cardiovascular disease (CVD) mortality prediction in Israeli industrial workers, and developed and validated new risk prediction models for CVD mortality incidence in the same population. Our database was a longitudinal Israeli industrial cohort (CORDIS cohort) of 4809 adult males followed-up for 22 years. Performance of the FHS and the SCORE prediction models was analyzed by insertion of the CORDIS cohort measurements to each model separately. The standard prognostic variables and results obtained from the new refined Cox regression analyses were used to construct two new 10- and 20-year CVD mortality risk scoring systems: a modified FHS model (FHS/Cox) and an omnibus model with Cox regression (Omnibus/Cox). The SCORE model of high-risk and low-risk charts yielded 10-year mortality mean risks of 1.12% and 0.64%, respectively, for male subjects aged > 30 years. The new FHS/Cox and Omnibus/Cox models generated a mean predictive 10-year risk of 1.12% and 1.50%, respectively. The mean 20-year risk predicted by the new FHS/Cox and the Omnibus/Cox models was 2.66% and 3.75%, respectively. Internal validation of both models demonstrated a high and stable area under the receiver operating characteristic curve > 0.85. No significant differences were found between the two models. In conclusion, the CVD mortality risk prediction scoring systems tailored for the Israeli workers population demonstrated good performance. Additional studies to externally validate these algorithms will indicate which of these quantitative risk estimation platforms should be used in specific settings.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

The concept of risk assessment and reduction, initially introduced by the American Framingham Heart Study (FHS) > 50 years ago and refined by other models, forms the cornerstone of preventive cardiology (Pearson, 2002; National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), 2002; Greenland et al., 2010; Perk et al., 2012). Risk factor assessment, the first step in primary cardiovascular disease (CVD) prevention, guides the therapeutic strategy as the intensity of preventive efforts is tailored to each patient’s unique CVD risk status (Pearson, 2002; National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III), 2002; Greenland et al., 2010; Perk et al., 2012).

Risk prediction algorithms have been developed and used to identify high-risk individuals. The most well-established risk score algorithms are the FHS risk scores (Gibbons et al., 2013; Dawber et al., 1963) and the European Systematic Coronary Risk Evaluation (SCORE) (Conroy et al., 2003). The SCORE risk scores were calculated for high- and low-risk regions of Europe, using a database including > 200,000 patients pooled from cohort studies in 12 European countries (Conroy et al., 2003). The Omnibus algorithm, published in the 2013 American College of Cardiology (ACC)/American Heart Association (AHA) Guidelines on the assessment of CVD risk, calculates the risk of a first atherosclerotic cardiovascular disease (ASCVD) event (Goff et al., 2014). This algorithm is based on a Cox regression model.

Several basic differences exist between the 3 models. The FHS risk scores estimate the 10-year risk of developing CHD (Gibbons et al., 2013; Dawber et al., 1963) while the SCORE, estimates the 10-year risk of a first fatal ASCVD event (i.e. CVD mortality) (Conroy et al., 2003). The Omnibus model estimates the 10-year risk of a first ASCVD event, defined as a nonfatal myocardial infarction or CHD death, or fatal or nonfatal stroke, among people free of ASCVD (Goff et al., 2014). Common prediction factors for the models include: age, gender, total cholesterol, high-density lipoprotein cholesterol (HDL-C), systolic blood pressure (SBP) and current smoking status. Other unique variables include “belonging to high- or low-risk regions of Europe” for...
Baseline characteristics of the Israeli industrial cohort followed-up for 22 years.

Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>CORDIS cohort N = 4809</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years</td>
<td>42.3 ± 12.1</td>
</tr>
<tr>
<td>Cholesterol (mg/dL)</td>
<td>202.1 ± 44.6</td>
</tr>
<tr>
<td>High-density lipoprotein cholesterol (mg/dL)</td>
<td>42.8 ± 11.3</td>
</tr>
<tr>
<td>Systolic blood pressure (mm Hg)</td>
<td>125.6 ± 16.0</td>
</tr>
<tr>
<td>Body mass index (kg/m²)</td>
<td>25.7 ± 3.2</td>
</tr>
<tr>
<td>Resting heart rate (beats/min)</td>
<td>76.6 ± 10.8</td>
</tr>
<tr>
<td>Smoker at baseline, N (%)</td>
<td>1900 (39.5)</td>
</tr>
<tr>
<td>Diabetes mellitus at baseline, N (%)</td>
<td>159 (3.3)</td>
</tr>
<tr>
<td>Treated for hypertension, N (%)</td>
<td>334 (6.9)</td>
</tr>
</tbody>
</table>

Continuous variables are displayed as mean ± standard deviation; categorical variables are displayed as N (%).

The Israeli population has specific characteristics in terms of lifestyle and genetics. Nonetheless, clinicians use the FHS and the SCORE risk score charts for predicting CHD risk in Israeli individuals, even though they have never been validated in this population.

Our primary objective was to examine the performance of these coronary mortality risk prediction models in the Israeli population and to develop new models better fitted to this population thus creating an adjusted and reliable quantitative risk estimation platform to be potentially used by local clinical decision-makers.

2. Methods

2.1. Study design/participants

We used the Cardiovascular Occupational Risk Factor Determination in Israel Study (CORDIS) population database to 1) analyze its performance in the FHS and SCORE risk prediction models, and 2) for the development of two new risk prediction models.

The CORDIS population included 7661 male and female workers aged 18–75, recruited from 21 industrial plants (metalwork, textiles, light industry, electronics, food manufacturing and plywood production) throughout Israel for on-site screening of cardiovascular risk factors. Approval for the study was obtained from the Ethics Committees of the National Institute of Occupational and Environmental Medicine and the Chaim Sheba Medical Center, Ramat Gan, Israel. All participants provided written informed consent.

The current analysis was restricted to 4809 male employees aged 20–75 years at baseline (excluded: <20 or ≥75 years, N = 68). Females (N = 2399) were excluded from the analysis due to the small proportion of CVD deaths reported (8 and 21 women died of CVD in the 10-year and 20-year time points, respectively). Arab men (N = 259) were excluded from the current analysis as no follow-up data was available for them; individuals who reported myocardial infarction at baseline were also excluded (N = 126) (Fig. 1).

2.2. Data collection

Data collection was performed as previously described (Harari et al., 2015). Briefly, data related to medical history, demographics, blood test results and cardiovascular risk factors were collected in two phases: 1985–1987 and 1988–1990. Trained technicians visited the various plants and performed computerized interviews as well as physical examinations. The CORDIS questionnaires constituted the basis for population characterization and the identification of risk factors.

Table 3

Mean 10- and 20-year risk of cardiovascular disease mortality in an Israeli industrial cohort followed-up for 22 years estimated according to the new models: FHS/Cox and Omnibus/Cox.

<table>
<thead>
<tr>
<th>Risk algorithm</th>
<th>Mean risk (%)</th>
<th>95% CI</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>FHS/Cox (10-year)</td>
<td>1.12</td>
<td>1.06, 1.18</td>
<td>0.01</td>
<td>45.49</td>
</tr>
<tr>
<td>Omnibus/Cox (10-year)</td>
<td>1.5</td>
<td>1.43, 1.56</td>
<td>0</td>
<td>33.09</td>
</tr>
<tr>
<td>FHS/Cox (20-year)</td>
<td>2.66</td>
<td>2.53, 2.8</td>
<td>0.03</td>
<td>77.75</td>
</tr>
<tr>
<td>Omnibus/Cox (20-year)</td>
<td>3.75</td>
<td>3.58, 3.92</td>
<td>0</td>
<td>65.09</td>
</tr>
</tbody>
</table>

N = 4809; FHS, Framingham Heart Study; CI, confidence interval.
Approximately 900 variables were collected for each participant, including workplace variables (Froom et al., 2004). The data obtained from the CORDIS cohort participants in both phases was merged in 2007, and updated in 2012 with mortality data obtained from the National Death Registry of the Israel Ministry of the Interior and the Central Bureau of Statistics. CVD mortality was defined by ICD-9 codes 390–398, 402, 404, 410–414, 429.2. The data for 693 male subjects (14.4%) could not be merged as they had left the country. These subjects were considered alive and included in the analyses.

2.3. Calculation of risk scores using established prediction models

Performance of the FHS and SCORE prediction models in the CORDIS population was analyzed by inserting the CORDIS population–based
cohort measurements to each model algorithm using the same parameters and coefficients originally included in FHS and SCORE prediction models. Although The FHS risk model is used for predicting CHD morbidity, we extrapolated it to predicting mortality taking into account that FHS scores will obviously tend to be higher than the actual risks of mortality. The FHS risk score for CVD morbidity is calculated by adding up the total risk scores for the following risk factors: age, HDL-C level, total cholesterol level, SBP (treated or untreated), smoking status (yes/no) and diabetes (yes/no). In contrast to the dataset used for the FHS risk model, HDL-C level and treated SBP in the CORDIS data set were not found to be significant by Cox regression; these variables were therefore excluded from the FHS algorithm used in the CORDIS cohort.

2.4. Construction of the FHS/Cox model based on CORDIS cohort characteristics

We developed a combined model according to the method described by Sullivan et al. (Sullivan et al., 2004), using the FHS method together with a Cox model. The sample used for the calculations and regression analysis consisted of 4809 males aged 20–75 years with no history of myocardial infarction. Participants who had left the follow-up were assigned an end-of-tracking date and Kaplan Meier survival analysis was done in order to take into account incomplete follow-up information. A sensitivity analysis in which all dropouts were removed produced similar results (data not shown).

Using the Cox proportional hazards model, regression coefficients and hazard ratios (calculated with a 95% confidence interval [CI]) were calculated for the following risk factors: age (years) (95% CI, 0.08439, 1.088); SBP (mmHg) (95% CI, 1.10638, 3.023), total cholesterol (gr/dL) (95% CI, 0.71076, 2.036), current smoking status (95% CI, 0.56576, 1.761) and diabetes status (95% CI, 0.87977, 2.410). Age, as a continuous variable, was divided into different categories and the average age of each category was determined as its reference value (i.e. each mid-point category). The remaining risk factors were modelled using sets of dummy variables. After selecting a reference risk factor profile (by choosing a category for each risk factor), the “distance” between each risk factor category and the reference category was calculated in terms of regression units. Each subject was assigned a total score following the addition of each of the 5 risk factor-assigned points. The sum of these points (an integer between − 1 and 14) gave the final value for either the 10-year risk score or the 20-year risk score for CVD mortality, separately.

2.5. Construction of the Omnibus/Cox model based on CORDIS cohort characteristics

A similar approach was adopted using multivariable risk equations and the Omnibus method for the prediction of the 10-year risk of ASCVD events, based on the Cox model using a logarithmic scale and the remaining significant interactions (Goff et al., 2014) (in our model no interactions remained). This method yielded a slightly different calculation of the 10- and 20-years risk of CVD mortality compared to our FHS/Cox model. However, this approach does not approximate the regression coefficients that are in the Framingham scoring system; rather it uses the exact risk equations of the model for predicting the risk of each patient according to his/her exact risk factor values.

We used the area under the receiver operating characteristic (AUROC) curve to assess the predictive value of the model. In addition, internal receiver operating characteristic (ROC) curve validation was performed by randomly splitting the sample to a 2000-subject subsamples, repeating this step 100 times and evaluating the performance of both models for 10- and 20-year risk prediction. Of note, the internal ROC validation procedure considers all data obtained for the 22 years of follow-up, even if the calculation is for 10 years. Therefore the internal ROC curve validation for the 10-year period considered all 170 CVD death events. The above internal ROC curve validations can be compared to reproducibility (Thomsena et al., 2002), as each random subsample serves as a subpopulation.

3. Results

3.1. General characteristics of the study population at baseline

A total of 4809 participants were included in the analysis. The median follow-up period was 22 years (mean follow-up time, 22.1 ± 3.2 years). The mean age of the study population at baseline was 42.3 ± 12.1 years (Table 1).

There were 76 newly-recorded CVD mortality events during the 10-year follow-up period, and 170 newly-recorded CVD mortality events during the 20-year follow-up period, corresponding to a crude incidence rate of 1.60 (95% CI, 1.27, 2.02) and 1.84 (95% CI, 1.58, 2.15), respectively, per 1000 person-years.

3.2. CORDIS data prediction of 10-year CVD morbidity/mortality risk based on the Framingham and EUROSCORE risk algorithms

We first examined the 10-year CVD risk scores obtained for 3836 and 3870 male subjects from the CORDIS population using the EUROSCORE and Framingham (Conroy et al., 2003; Crimmins & Beltrán-Sánchez, 2011) risk algorithms, respectively. In accordance with the European SCORE and FHS guidelines, for this analysis, we excluded subjects aged 20–30 years (for both models) and subjects aged 70–75 years (for the European SCORE model) as these age groups were not included in the European SCORE and FHS risk score models (Perk et al., 2012; Conroy et al., 2003; Goff et al., 2014). Since the SCORE risk estimation system was developed for use in both high- and low-risk asymptomatic persons, both risk charts were applied to the CORDIS data. As shown in Table 2, the mean predictive 10-year CVD mortality risks obtained by applying the SCORE high-risk and low-risk charts to CORDIS male subjects...
were 1.12% (95% CI, 1.05, 1.18) and 0.64% (95% CI, 0.60, 0.68), respectively.

The original FHS risk score system generated an estimated risk of 10.7% (95% CI, 10.43, 10.98). As this system was developed to estimate CVD morbidity risk rather than mortality, an estimated risk of 10.7% is therefore plausible.

3.3. Estimations of 10- and 20-year CVD mortality risk in Israeli industrial workers using the novel study model

Based on the coefficients obtained from the Cox regression analysis, we used the prognostic variables age, SBP, cholesterol, diabetes and smoking status to construct two risk score algorithms: a modified FHS model (FHS/Cox) and an Omnibus model (Omnibus/Cox) with Cox regression, in order to predict the 10- and 20-year CVD mortality risk using the CORDIS population data set. As demonstrated in Table 3, the mean predictive 10-year risk generated by the FHS/Cox model and the Omnibus/Cox model was 1.12% (95% CI, 1.06, 1.18) and 1.50% (95% CI, 1.43, 1.56), respectively. The mean 20-year risk predicted by the FHS/Cox model and the Omnibus/Cox model was 2.66% (95% CI, 2.53, 2.8) and 3.75% (95% CI, 3.58, 3.92), respectively.

3.4. Validation of the discrimination ability of the new models by AUROC curve

In order to assess the discrimination ability of the models, ROC curves were generated for the SCORE high/low-risk (Fig. 2, panels A–B), the 10- and 20-year risk FHS/Cox model (Fig. 2, panels C–D) and the 10- and 20-year risk Omnibus/Cox model (Fig. 2, panels E–F).

The AUROC curve obtained was 0.815, 0.793, 0.847, 0.854, 0.849 and 0.857 in each model, respectively. It should be noted that the ROC curve for the Omnibus/Cox model is “serrated” compared to the FHS/Cox ROC curve due to an absolute different risk calculated and plotted for each CORDIS subject, as opposed to when using the FHS, in which categorical scores are assigned to each subject, resulting in a higher proportion of identical scores.

Internal ROC curve validations were performed by randomly splitting the entire dataset to a 2000-subject subsamples, repeating this step 100 times and evaluating its performance separately for 10- and 20-year risk prediction using each model. The AUROC results, as measures of prediction accuracy, are summarized in Fig. 3. The mean AUROC was high in all models: 0.868 and 0.877 for the 10- and 20-year FHS/Cox risk model, respectively; and 0.880 and 0.884 for the 10- and 20-year Omnibus/Cox risk model, respectively.

**Fig. 4.** Predicted probabilities of death versus calculated risks of 10- and 20-year risk of CVD mortality according to FHS/Cox (A, B), Omnibus/Cox (C, D) and SCORE (E, F) models in an Israeli industrial worker-based cohort followed-up for 22 years. The predicted probabilities of CVD mortality (with bounding 95% confidence limits, grey area) calculated by a logistic model with the actual date of death due to CVD of all subjects (N = 4809), were plotted against the calculated 10- and 20 year risk of CVD mortality according to the relevant model.
The predicted probabilities of CVD mortality, calculated by a logistic model with the actual date of death due to CVD of all subjects (N = 4809), were plotted against the calculated 10- and 20 year risk of CVD mortality according to the relevant model (Fig. 4). It seems that the model that best-fit the predicted probability values of CVD mortality calculated by the actual death data, is the 20-year risk Omnibus/Cox model, as can be seen by the narrow confidence interval area in Fig. 4D.

4. Discussion

The FHS risk score is considered the “gold standard” in clinical practice for CHD risk assessment (Pearson, 2002) and is used as a tool to predict the 10-year risk of CVD morbidity in individuals aged 30–74 years with no history of CVD at their baseline examination (Sullivan et al., 2004). Risk estimation is based on group averages that are then applied by the clinician to individual patients (Goff et al., 2014). A number of persistent concerns with existing risk equations have been noted, including non-representative or historically-dated populations, limited ethnic diversity, narrowly-defined endpoints, endpoints influenced by provider preferences and endpoints with poor reliability (Ramsay et al., 2011; Hemann et al., 2007; Thomsena et al., 2002). Nonetheless, two systematic reviews based on several studies support the conclusion that risk assessment, combined with counseling is associated with favorable but modest changes in patient knowledge and intention to modify risk-related habits, and with provider-prescribed behavior and risk factor control (Sheridan & Crespo, 2008; Sheridan et al., 2010).

In this report we present two new risk algorithms for the assessment of CVD mortality risk in the Israeli industrial worker population. Similar to the SCORE project, the current analysis calculated the risk of mortality rather than morbidity. In an attempt to assess the utility of long-term and lifetime risk assessment as an adjunct to short-term (10-year) risk assessment, we also examined the 20-year risk, in both models. Extensive epidemiological, pathological, and basic science data indicate that the development of atherosclerosis, the precursor of ASCVD, occurs over a timespan of decades and is related to long-term and cumulative exposure to causal, modifiable risk factors (Jamison et al., 2006). Thus, a life-course perspective to risk assessment and prevention must be considered and primary importance should be given to the potential benefits of lifestyle modification (Goff et al., 2014).

The original FHS risk score system generated an estimated risk of 10.7% for the CORDIS cohort. As this system was developed to estimate CHD morbidity risk rather than mortality, an estimated risk of 10.7% is plausible.

The mean predictive 10-year risks obtained by applying the high-risk and low-risk SCORE charts were 1.12% and 0.64%, respectively. The use of the SCORE low-risk algorithm may be suboptimal, implying a significant effect of environmental risk factors on the Israeli population. Interestingly, the calculated high-risk SCORE was identical to the value obtained in our novel FHS/Cox model in contrast to the recommendation made by the European Guidelines on CVD prevention in clinical practice (2012) (Perk et al., 2012) to use the low-risk charts in the Israeli population. In the case of the CORDIS cohort, most of the workers were blue collar workers from a lower socioeconomic background. Increased mortality risk has been associated with lower socioeconomic background (Holme et al., 1980; Marmot et al., 1984), supporting the use of the high-risk European SCORE charts in assessing CVD risk in asymptomatic industrial workers in Israel. However, it remains to be investigated if this finding should be applied to the entire male Israeli population.

In this study we developed two novel risk algorithms: the first is based on a modified FHS risk algorithm, and the second is based on the Omnibus algorithm with Cox regression. Using these unique algorithms we were able to predict the 10- and 20-year risk of CVD mortality in a relatively large sample of Israeli industrial workers that is homogeneous in terms of socioeconomic status and ethnic background. Internal validation of both models demonstrated a high and stable AUROC >0.85, indicating good predictive power as well as accuracy and generalizability of the new prognostic models. No significant differences were found between the two models. Furthermore, our results indicate that the FHS/Cox scoring approximation does not affect the exact predictive values as in the Omnibus/Cox method and that the Omnibus/Cox logarithmic scale does not improve the basic Cox model as in the FHS method, concluding that the FHS/Cox may be easier to use.

The CORDIS database is large, up-to-date relatively complete and accurate. It is therefore an appropriate database for the construction of a risk evaluation scoring system for use in Israeli male industrial workers. Nevertheless, the study has several limitations. We were unable to introduce large sectors of the entire industrial population to the scoring systems: woman and Arabs (due to low mortality or missing data) and had a relatively narrow age range as the majority of subjects (>90%) were between 30 and 65 years of age. These limitations make it impossible to extend the application of the models to the entire Israeli population, and thus the claims of the model are weakened. Given the lack of external validation for this study, the generalizability of the findings to persons of other ethnic backgrounds or Israeli non-industrial workers will need further testing. Furthermore, possible bias related to the accuracy of the primary reason for death registered at the National Death Registry as well as the fact that subjects who had left the country were considered still living might have led to a limited underestimation of the calculated risk scores.

The main strength of the present study lies in its relatively large sample size and long duration of follow up (22 years). Second, the use of individual participant data enabled a consistent approach to adjustment for potential confounders and the exploration of additional risk factors for improvement of the predictive validity of the models. Although there was variability in the risk score estimations by the two models, its extent was not greater than that seen with other similar studies, and may be partly attributed to the data collection method.

In conclusion, to the best of our knowledge, the risk scoring systems presented here are the first of their kind to be tailored to the Israeli population. External validation of these algorithms will enable to identify individuals at risk of CVD and to recommend which of these quantitative risk estimation platforms should be used in specific settings or locations by local health care providers and decision makers.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare there is no conflict of interest.

References


