Poor predictive ability of the risk chart SCORE in a Danish population

Madina Saidj[1], Torben Jørgensen[1, 2, 3], Eva Prescott[4] & Anders Borglykke[1]

ABSTRACT

INTRODUCTION: In Denmark, the European risk chart Systematic COronary Risk Evaluation (SCORE) from the European Society of Cardiology is recommended for use in cardiovascular prevention. Nevertheless, its predictive ability in a Danish population has never been investigated. The purpose of this study was therefore to assess the predictive ability of the SCORE risk chart with regard to fatal cardiovascular risk according to the socio-demographic factors of age, sex, income and education in a Danish population.

MATERIAL AND METHODS: Data from the third Copenhagen City Heart Study (n = 4,224) were linked to the Danish Cause of Death Registry. Calibration (i.e. Hosmer-Lemeshow goodness-of-fit), expected-to-observed (E/O) mortality ratios in the total population and for subgroups, as well as discrimination (i.e. sensitivity, specificity, area under the Receiver Operator Characteristic (AUROC) and predictive values) were tested. Both SCORE high-risk and low-risk were applied for comparison.

RESULTS: The results showed that both SCORE high-risk and low-risk performed acceptably in terms of discrimination (AUROC ≈ 0.7-0.8); however, calibration for both SCORE charts was inadequate ($\chi^2 > 20$; $p < 0.001$). E/O-ratios varied with age, sex and socioeconomic status.

CONCLUSION: There is a need to recalibrate SCORE to risk levels and risk factor distribution in the Danish population.

FUNDING: not relevant.

TRIAL REGISTRATION: not relevant.

Cardiovascular disease (CVD) remains one of the leading causes of mortality in developed countries and it is the number one cause of death globally [1]. Comprehensive cardiovascular risk assessment models have consequently been developed to assist prevention and ensure early identification and treatment of persons at high risk of dying from CVD. The European risk chart Systematic COronary Risk Evaluation (SCORE) is such a risk assessment model.

It is derived from data from > 200,000 individuals pooled from 12 European cohort studies collected in the 1980-90s (from Denmark: several population studies from Research Centre for Prevention and Health), and constructed by the European Society of Cardiology (ESC) with the aim of providing better predictive ability for European individuals. SCORE estimates the ten-year risk of a fatal CVD event for individuals in the age range 40-65 years without diabetes [2].

The ESC guidelines for CVD prevention in clinical practice advocate the use of SCORE for more rigorous identification of asymptomatic individuals who are at increased risk of CVD [3]. The Danish Society of Cardiology has embraced these guidelines, and since 2004 the systematic use of SCORE in primary practice has been recommended in Danish guidelines on CVD prevention [4].

Like in many other European populations, cardiovascular mortality has declined in Denmark during recent decades [5]. Risk estimates based on cohort studies that started more than a decade ago are thus likely to estimate cardiovascular risk incorrectly. The applicability of SCORE has been evaluated in several European countries. Studies show that SCORE generally overestimates the CVD risk [6-10], and so the chart has been recalibrated for a number of European countries [11]. Despite national guidelines dating back to 2004 advising systematic use of the risk chart in general practice in Denmark [4], the predictive ability of SCORE in a Danish population has never been investigated.

The individual SCORE risk estimation is based on gender, age, smoking status, systolic blood pressure, and total cholesterol. A series of other CVD risk factors are classified as 'qualifiers'. Social deprivation is such a qualifier [3, 12]. This means that risk may be higher than indicated by SCORE in socially deprived individuals. Yet, evidence is limited on the performance of SCORE in different socioeconomic groups.

The aim of this study, therefore, was to use population-based data to assess the predictive ability of SCORE in Denmark by the socioeconomic factors of income and education.

MATERIAL AND METHODS

Copenhagen City Heart Study

The Copenhagen City Heart Study (CCHS) is a prospective cardiovascular study of 20,000 women and men aged 20 years and older, randomly drawn from the Copenhagen Population Register [13]. Four studies were conducted (1976-78, 1981-83, 1991-94, 2001-2003). The third study (1991-1994) of the CCHS was selected for this study because it is the most recent study for which the prerequisite 10-year follow-up was available from

ORIGINAL ARTICLE

1) Research Centre for Prevention and Health, Glostrup Hospital
2) University of Copenhagen
3) University of Aalborg
4) Department of Cardiology, Copenhagen City Heart Study, Bispebjerg Hospital

Dan Med J 2013;60(5):A4609

The Danish Register of Cause of Death (data were available up to 31-12-2009). Information on examination procedures and distribution of risk factors is described in Schnohr et al, 2001 [13].

Study sample

From a total study sample of 10,135 individuals from the third study by the CCCHS, 4,224 subjects were included in the analyses. Reasons for exclusion were: failure to fulfil the age restriction of SCORE (40-65 years) (n = 5,441, 54%), previous CVD events (n = 229, 2.3%), diabetes (n = 131, 1.3%) and missing values on covariates (n = 110, 1.1%).

Definition of endpoints

The study population was traced by means of the unique identification number of all Danish citizens which was linked to the Danish Cause of Death Registry. The ICD codes of death certificates used to create SCORE [2] were applied, i.e. International Classification of Diseases (ICD)-10 codes I10-I25, I44-I51, I61-I73 and R96-R96.1 and the corresponding ICD-8 codes 401-414, 426-443 (with the exception of 426.7, 429, 430, 432.1, 437.3, 437.4, 437.5) and 798.1, 798.2

Statistical analyses

The predictive ability of SCORE was studied by means of calibration measures, expected-to-observed (E/O) mortality ratios and discrimination measures. Calibration is the extent to which predicted and observed events coincide. For this the Hosmer-Lemeshow goodness-of-fit statistic was used. Low values indicate good calibration, and values > 20 indicate significant lack of calibration ($p < 0.01$) [14]. The E/O-ratios supplemented the calibration by assessing the (dis)agreement in subgroups of age, sex and socioeconomic indicators. Discrimination is the model's ability to correctly rank individuals according to risk. For this sensitivity, specificity and predictive values at the risk threshold of SCORE (≥ 5%) as well as the area under the receiver operator characteristic (AUROC) curve was assessed. The area under the curve is the probability that a person who experiences an event will have a higher risk score than a person who does not experience an event.

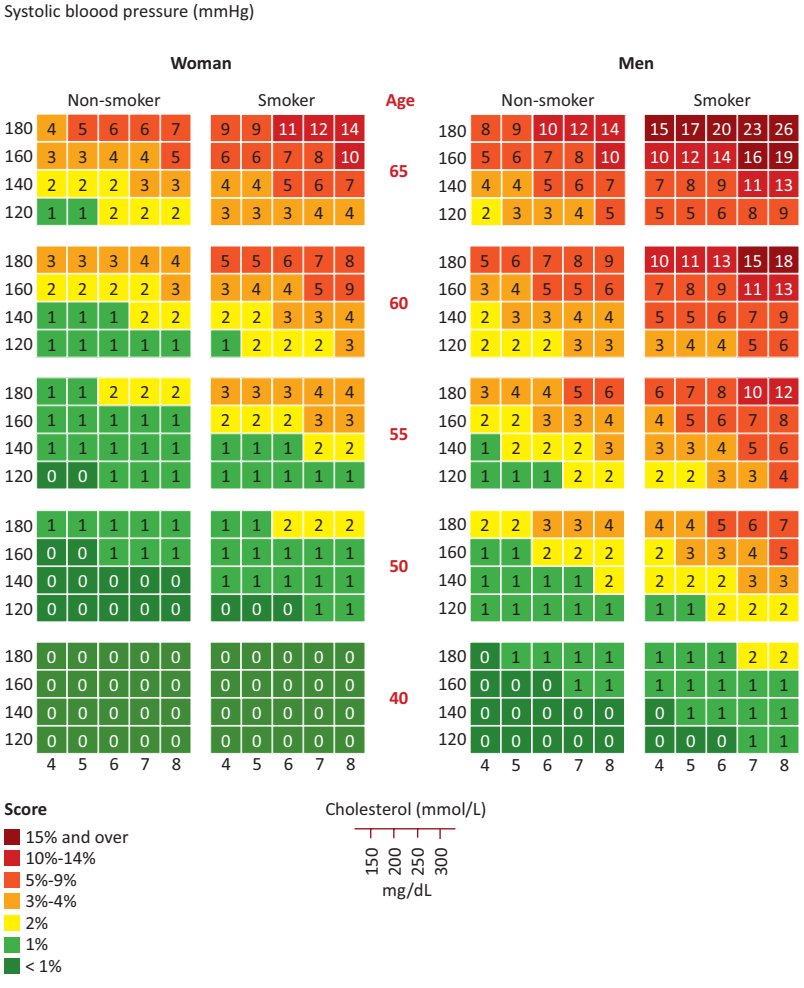
In 2012 Denmark was re-categorized from a high CVD risk region to a low CVD risk region and hence now uses the low-risk version of SCORE (**Figure 1**). Consequently, the number of expected CVD events estimated by both SCORE high-risk and low-risk was calculated for comparison. This calculation used the same methodology, i.e. risk factors (age at baseline, sex, smoking yes/no, systolic blood pressure (mmHg) (measured in a sitting position on the left upper arm after five minutes of rest, determined by a London School of Hygiene sphygmomanometer [13]) and total cholesterol (mmol/l) (determined by cholesterol oxidase-phenol-aminophenazone (CHOD-PAP) method [13]) were used) and risk coefficients as in SCORE [2]. Socioeconomic position was defined by income, measured as the total gross household income (categorized into low (< 100,000 DKK), lower middle (100,000-199,000 DKK), upper middle (200,000-399,000 DKK) and high (≥ 400,000 DKK)) and education measured as vocational training (categorized into none, ≤ 1 year, 1-3 years, > 3 years). The expected number of CVD events was calculated as the sum of individual absolute risks [6].

The statistical analyses were performed using SAS version 9.2. Data are presented as means ± standard deviations and percentage frequencies.

Trial registration: not relevant.

FIGURE 1

SCORE (Systematic COronary Risk Evaluation) – European low-risk chart. Ten-year risk of fatal cardiovascular disease in low-risk regions of Europe by gender, age, systolic blood pressure, total cholesterol and smoking status.



RESULTS

The baseline characteristics of the study population by sex, age, income and education are given in **Table 1**. Of the 4,224 subjects in our study population, 157 (114 men and 43 women) developed a fatal cardiovascular event over the course of ten years. The number of events predicted by SCORE was 212 for the high-risk model and 122 for the low-risk model. When looking at the distribution of the actual, observed events by the SCORE risk classification of the study population (**Table 2**), there was good agreement for men: the majority experiencing a fatal CVD event were classified by SCORE as high-risk subjects (86.8% and 70.2% of all events, respectively, for SCORE high-risk and SCORE low-risk). The agreement was less satisfactory for women: of the 43 women experiencing an event, 23 (53.5 %) and 27 (62.8%), respectively, were classified in the intermediate risk of 1-4%.

Hosmer-Lemeshow χ^2 statistic was ± 33 ($p < 0.01$), exceeding the limit of 20, and thereby indicating significant lack-of-fit for both models. However, when stratifying by sex, both models showed a better fit for women, slightly better for SCORE low-risk ($p \approx 0.19$, $\chi^2 \approx 11.2$) than for SCORE high-risk ($p \approx 0.13$, $\chi^2 \approx 12.4$). For men both models still lacked fit with the data, with SCORE low-risk ($p < 0.01$, $\chi^2 \approx 34.4$) performing slightly poorer than SCORE high-risk ($p \approx 0.01$, $\chi^2 \approx 31.2$).

In **Table 3** the E/O ratios are stratified by age, sex, income and education. Overall, SCORE high-risk overestimated the number of events, especially among women. SCORE low-risk was a relatively good predictor among women and among men in the age range 45-54 years, but underestimated the risk among men aged ≥ 55 years. SCORE high-risk was a better predictor among the lower income and education groups, while SCORE low-risk was a better predictor among the higher income and education groups. This was applicable for both women and men – though most markedly for men.

In **Table 4**, we quantify and compare the capacity of the models to predict events by means of the ROC curve and by comparing sensitivities, specificities and predictive values at the recommended risk threshold for intervention (≥ 5%). With areas under the ROC-curve of ≈ 0.7-0.8 for both models, the discriminative predictive ability of SCORE in the study population was acceptable. Sensitivity was best among men, and, conversely, specificity was best among women. The 5% threshold showed good balance between sensitivity and specificity. The balance was better among men than among women.

TABLE 1

Baseline characteristics of the study population.

	All	Men	Women
n	4,224	1,900	2,324
Age, % (n)			
40-44 yrs	13.5 (571)	13.3 (253)	13.7 (318)
45-49 yrs	14.6 (615)	16.2 (308)	13.2 (307)
50-54 yrs	21.0 (886)	22.4 (425)	19.8 (461)
55-59 yrs	22.6 (954)	21.6 (411)	23.4 (543)
60-65 yrs	28.4 (1,198)	26.5 (503)	29.9 (695)
Age, yrs, mean (± SD)	54.1 (± 7.2)	53.8 (± 7.1)	54.4 (± 7.2)
Smokers, % (n)	55.4 (2,341)	47.5 (1,111)	52.5 (1,230)
SBP, mmHg, mean (± SD)	135.4 (± 20.1)	138.4 (± 19.4)	132.9 (± 20.3)
TC, mmol/l, mean (± SD)	6.2 (± 1.2)	6.1 (± 1.1)	6.3 (± 1.3)
Income[a], % (n)			
Low	9.9 (407)	9.2 (172)	10.5 (235)
Lower middle	25.7 (1,059)	22.0 (412)	28.8 (647)
Upper middle	40.6 (1,671)	41.2 (779)	39.7 (892)
High	23.8 (979)	27.2 (509)	20.9 (470)
Education[b], % (n)			
None	20.3 (855)	14.9 (282)	24.8 (573)
≤ 1 yr	29.3 (1,231)	43.2 (816)	18.0 (416)
1-3 yrs	30.7 (1,289)	18.7 (354)	40.4 (935)
> 3 yrs	19.7 (826)	23.1 (436)	16.8 (390)

SBP = systolic blood pressure; SD = standard deviation; TC = total cholesterol concentration.
a) Income is measured as the total gross household income for the calendar year preceding the year of data collection, N = 4,116 (1,876 men, 2,244 women).
b) Education is measured as vocational training, N = 4,202 (1,888 men, 2,314 women).

TABLE 2

Distribution of observed fatal cardiovascular disease events by Systematic COronary Risk Evaluation (SCORE) risk classification. The values are % (n).

	SCORE, high-risk			SCORE, low-risk		
	all	men	women	all	men	women
High risk: ≥ 5%	75.8 (119)	86.8 (99)	46.5 (20)	59.2 (93)	70.2 (80)	30.2 (13)
Intermediate risk: 1-4%	23.6 (37)	12.3 (14)	53.5 (23)	36.3 (57)	26.3 (30)	62.8 (27)
Not at risk: < 1%	0.6 (1)	0.9 (1)	–	4.5 (7)	3.5 (4)	7.0 (3)

DISCUSSION

Overall, we found that SCORE has good discriminative ability. However, calibration is inadequate and E/O ratios vary with age, sex and socioeconomic status in this Danish population. This is in line with the common observation that SCORE performs relatively well in different populations in terms of discrimination, while calibration varies widely [2, 8]. SCORE high-risk, which until 2012 was recommended in Denmark, substantially overestimated CVD events in the population. This finding is in accordance with observations in other Nordic countries [6, 7, 9]. Besides being a better predictor in women, SCORE low-risk, which is today's recommended version, did not perform much better when predicting absolute risks.

The discriminative values match those previously reported in the literature (AUROC = 0.71-0.84) [15]. Few studies have reported sensitivity and specificity, but for those which have, our findings are comparable [8, 9]. A

TABLE 3

Observed and expected number of fatal cardiovascular disease events within ten years by age, income and vocational training.

	Observed events			SCORE high-risk expected events			E/O ratio			SCORE low-risk expected events			E/O ratio		
	all	men	women	all	men	women	all	men	women	all	men	women	all	men	women
Total	157	114	43	211.9	145.6	66.2	1.35	1.28	1.54	122.2	79.1	43.1	0.78	0.67	1.00
Age															
40-44 yrs	5	5	0	4.0	3.5	0.6	0.80	0.70	–	2.1	1.7	0.3	0.42	0.34	–
45-49 yrs	5	5	0	10.0	8.4	1.6	2.00	1.68	–	5.2	4.3	1.0	1.04	0.86	–
50-54 yrs	16	12	4	28.0	21.8	6.2	1.75	1.82	1.55	15.3	11.8	3.9	0.96	0.98	0.97
55-59 yrs	36	28	8	52.1	36.9	15.2	1.45	1.32	1.90	29.5	19.8	9.7	0.82	0.71	1.21
60-65 yrs	95	64	31	117.8	75.0	42.7	1.24	1.17	1.38	70.1	41.9	28.1	0.74	0.65	0.91
Income															
Low	30	21	9	27.9	16.5	10.7	0.91	0.78	1.19	16.1	9.0	7.0	0.54	0.43	0.78
Lower middle	43	31	12	63.6	40.5	23.1	1.48	1.31	1.92	37.5	22.4	15.1	0.87	0.72	1.26
Upper middle	61	46	15	83.2	60.1	23.2	1.36	1.31	1.55	47.5	32.5	15.0	0.79	0.71	1.00
High	20	14	6	32.6	25.7	6.9	1.63	1.79	1.15	18.2	13.7	4.4	0.91	0.98	0.73
For income[a]	154	112	42	206.8	142.8	64.0	1.34	1.28	1.53	119.2	77.6	41.5	0.78	0.69	0.99
Education															
None	40	24	16	47.5	24.5	23.0	1.19	1.02	1.44	28.3	13.3	15.0	0.71	0.55	0.94
≤ 1 yr	69	62	7	84.3	71.8	12.5	1.22	1.16	1.79	47.3	39.2	8.1	0.68	0.63	1.16
1-3 yrs	33	17	16	52.4	26.9	25.5	1.59	1.58	1.59	31.1	14.5	16.6	0.94	0.85	1.04
> 3 yrs	13	10	3	26.8	21.9	4.9	2.06	2.19	1.63	14.9	11.8	3.2	1.15	1.18	1.07
For education[a]	155	113	42	211.0	145.1	66.0	1.36	1.28	1.57	121.7	78.8	43.0	0.78	0.70	1.02

E/O = expected events/observed events.

a) For income (n = 4,116) and education (n = 4,202) a few cases were missing cf. Table 1. Therefore the ratios for these variables are presented separately.

lower sensitivity among women (and conversely higher specificity) is also in line with previous studies [16]. The low positive predictive values reflect the low prevalence of fatal CVD events in the study population. Regarding calibration, the performance of SCORE is obviously affected by the secular changes in CVD mortality between the time when the function was derived and the time at which it was applied.

Our analyses also point to substantial variation in CVD risk according to sex and age, and show that this variation exceeds what may be accounted for by SCORE. This result is in line with an Icelandic study of SCORE. In this study, the risk for younger men tallied with that in low-risk European countries and diverged towards that in high-risk countries with increasing age, while the risk for women was identical to that for women in low-risk countries [9]. Our analyses also showed that the predictive ability of SCORE varied by socioeconomic position. SCORE high-risk was a better predictor of risk in lower socioeconomic groups, while SCORE low-risk was a better predictor in higher socioeconomic groups. This finding is similar to the outcomes reported in a recent study in British men which found that overestimation by SCORE was particularly marked in high socioeconomic classes [17]. This finding also supports the classification of social deprivation as a SCORE qualifier [3] , and is in accordance with Danish literature on socioeconomic differences in CVD [18].

Some limitations of this study need to be addressed, one of which is that the study data date back to 1991-1994. This is an inherent limitation to any such study since ten-year follow-up is required. Another limi-

TABLE 4

Sensitivity, specificity, positive and negative predictive values at the 5% risk threshold, and AUROC for both models of SCORE.

	SCORE, high-risk					SCORE, low-risk				
	sensitivity	specificity	PPV	NPV	AUROC (95% CI)	sensitivity	specificity	PPV	NPV	AUROC (95% CI)
All	0.76	0.68	0.08	0.99	0.798 (0.765-0.831)	0.59	0.84	0.13	0.98	0.797 (0.765-0.830)
Men	0.87	0.49	0.10	0.98	0.764 (0.721-0.807)	0.70	0.73	0.14	0.97	0.765 (0.722-0.808)
Women	0.46	0.83	0.05	0.99	0.773 (0.716-0.830)	0.30	0.93	0.07	0.97	0.774 (0.718-0.831)

AUROC = area under the receiver operator characteristic; CI = confidence interval; NPV = negative predictive value; PPV = positive predictive value.

tation is the relatively low number of events, particularly among women (n = 44), which needs to be noted for reliability interpretation purposes, especially in age-stratified subgroups. A main strength of the study is the quality of the data used, both from the Danish Register of Cause of Death and the CCCHS.

The purpose of this study was to assess the predictive ability of SCORE in a Danish population. This study demonstrated that there is a need for recalibration of SCORE to risk levels and risk factor distribution in the Danish population. Because of secular time changes in CVD risk, a viable solution would be a continuous update of SCORE, which is feasible in Denmark where recent mortality statistics and epidemiological studies of risk factor distribution are accessible. What truly matters when assessing risk prediction models is the performance close to the levels at which decisions are made, not at the extremes when decisions are obvious. SCORE's classification of women at risk in the study population is therefore disappointing. This result adds to the growing evidence in favour of lowering the thresholds in female populations [16, 19] and amplifies the gain of a recalibration.

While estimating total risk seems eminently logical, it remains unknown whether an individual risk estimation approach results in risk factor – and ultimately morbidity – reduction [12], and research is warranted regarding clinical benefits and cost effectiveness. An innate consideration is the number needed to treat to achieve desirable outcomes. Finally, an important challenge for the future will be to introduce efficient public health policies with an emphasis on a population approach to support the current high-risk approach for cardiovascular prevention [20].

CORRESPONDENCE: *Madina Saidj*, Forskningscenter for Forebyggelse og Sundhed, Koncern Plan Udvikling og Kvalitet, Afsnit 84/85, Glostrup Hospital, 2600 Glostrup, Denmark. E-mail: m@saidj.dk

ACCEPTED: 14 February 2013

CONFLICTS OF INTEREST: Disclosure forms provided by the authors are available with the full text of this article at www.danmedbul.dk.

LITERATURE

1. Yusuf S, Reddy S, Ounpuu S et al. Global burden of cardiovascular diseases: part I: general considerations, the epidemiologic transition, risk factors, and impact of urbanization. Circulation 2001;24:2746-53.
2. Conroy RM, Pyörälä K, Fitzerland AP et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003;24:987-1003.
3. Perk J, De Backer G, Gohlke H et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012): The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). Eur Heart J, 2012 May 3 (epub ahead of print).
4. Thomsen T, Christensen B, Hildebrandt P et al. Kliniske retningslinier for forebyggelse af kardiovaskulær sygdom i Danmark. Tillæg til Cardiologisk Forum, 2004.
5. Danish Heart Association. Facts about cardiovascular diseases. Copenhagen: Danish Heart Association, 2012.
6. Lindman AS, Veierød MB, Pedersen JI et al. The ability of the SCORE high-risk model to predict 10-year cardiovascular disease mortality in Norway. Eur J Cardiovasc Prev Rehabil 2007;14:501-7.
7. Stenlund H, Lonneberg G, Jenkins P et al. Fewer deaths from cardiovascular disease than expected from the Systematic Coronary Risk Evaluation chart in a Swedish population. Eur J Cardiovasc Prev Rehabil 2009;16:321-4.
8. Ulmer H, Kollerits B, Kelleher C et al. Predictive accuracy of the SCORE risk function for cardiovascular disease in clinical practice: a prospective evaluation of 44 649 Austrian men and women. Eur J Cardiovasc Prev Rehabil 2005;12:433-41.
9. Aspelund T, Thorgeirsson G, Sigurdsson G et al. Estimation of 10-year risk of fatal cardiovascular disease and coronary heart disease in Iceland with results comparable with those of the Systematic Coronary Risk Evaluation project. J Cardiovasc Risk 2007;14:761-8.
10. Scheltens T, Verschuren WM, Boshuizen HC et al. Estimation of cardiovascular risk: a comparison between the Framingham and the SCORE model in people under 60 years of age. Eur J Cardiovasc Prev Rehabil 2008;15:562-6.
11. Cooney MT, Cooney HC, Dudina AL et al. Assessment of cardiovascular risk. Curr Hypertens Rep 2010;12:384-93.
12. Cooney MT, Dudina AL, Graham IM. Value and limitations of existing scores for the assessment of cardiovascular risk. J Am Coll Cardiol 2009;54:1209-27.
13. Schnohr P, Jensen G, Lange P et al. The Copenhagen City Heart Study. Tables with data from the third examination 1991-1994. Eur Heart J 2001;3(suppl H):H1-H83.
14. D'Agostino RB Sr., Grundy S, Sullivan LM et al. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180-7.
15. Berger JS, Jordan CO, Llyod-Jones D et al. Screening for cardiovascular risk in asymptomatic patients. J Am Coll Cardiol 2010;55:1169-77.
16. Bhalotra S, Ruwe MBM, Strickler GK et al. Disparities in utilization of coronary artery disease treatment by gender, race, and ethnicity: opportunities for prevention. J National Black Nurs Assoc 2007;18:36-49.
17. Ramsay SE, Morris RW, Whincup PH et al. Prediction of coronary heart disease risk by Framingham and SCORE risk assessments varies by socioeconomic position: results from at study in British men. Eur J Cardiovasc Prev Rehab 2011;18:186-93.
18. Danish Heart Association & NIPH. Danish Heart Statistics 2010. Copenhagen: Danish Heart Association and the Danish National Institute of Public Health (NIPH), University of Southern Denmark, 2010.
19. Ketola E, Laatikainen T, Vartiainen E. Evaluating risk for cardiovascular disease – vain or value? How do different cardiovascular risk scores act in real life? Eur J Public Health 2010;20:107-12.
20. Jørgensen T, Capewell S, Prescott E et al. Population level changes to promote cardiovascular health. Eur J Prev Cardiol, 2012 May 9 (epub ahead of print).