

Charles University in Prague, 1<sup>st</sup> Faculty of Medicine

**Cardiovascular Disease Risk Estimations  
Based on Data from Epidemiological  
Studies**

*Ph.D. Thesis Summary*

Prague 2006

**Author:** RNDr. Jindra Reissigová

**Address:** EuroMISE Centre

Institute of Computer Science AS CR

Pod Vodárenskou věží 2

182 07 Prague 8

tel.: +420-266 053 098

email: reissigova@euromise.cz

**Commission:** Biomedical Informatics

**Supervisor:** Prof. RNDr. Jana Zvárová, DrSc.



# Contents

<b>Preface</b>	<b>4</b>
<b>1 Introduction</b>	<b>5</b>
<b>2 Aims of the thesis</b>	<b>8</b>
<b>3 Material and methods</b>	<b>10</b>
3.1 Design of the STULONG study . . . . .	10
3.2 Validation studies within the STULONG study . . . . .	13
3.3 Statistical methods . . . . .	14
<b>4 Results</b>	<b>16</b>
4.1 Cardiovascular risk factors in the Czech Republic . . . . .	16
4.2 Validation of cardiovascular risk estimations in the Czech Republic . . . . .	17
4.2.1 Framingham risk function (1991) . . . . .	17
4.2.2 Framingham risk function (1998) . . . . .	20
4.2.3 SCORE risk function (2003) . . . . .	24
<b>5 Discussion</b>	<b>27</b>
5.1 Validation studies in the Czech Republic . . . . .	27
5.2 Other validation studies . . . . .	29
<b>6 Conclusions</b>	<b>31</b>
<b>References</b>	<b>32</b>
<b>Appendix: Publications and presentations of the author</b>	<b>35</b>

## Preface

The thesis *Cardiovascular disease risk estimations based on data from epidemiological studies* summarizes what has been published on estimations of a cardiovascular risk, and explores the validity of cardiovascular risk estimations in the Czech population. The text of the thesis is mainly comprised of the published papers in which I participated. These publications are listed in Appendix and cited in References.

I acknowledge that the thesis could not have been written without help and support my supervisor Prof. RNDr. Jana Zvárová, DrSc., who I thank for that. I also thank to Prof. MUDr. František Boudík, DrSc. and MUDr. Marie Tomečková, CSc., with whom I collaborated on the publications cited here. Finally, I thank for the support to institutional research plan AV0Z103000504 of the Institute of Computer Science the Academy of Sciences of the Czech Republic, the project 1ET200300413 of Academy Science of the Czech Republic, and the grant LN00B107 of the Ministry of Education of the Czech Republic.

RNDr. Jindra Reissigová

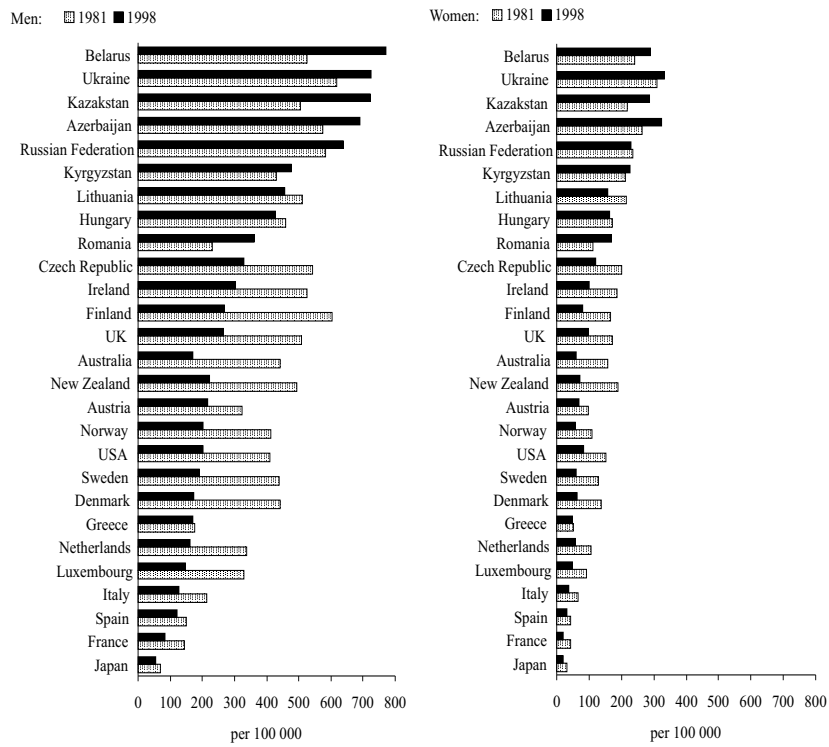
# 1 Introduction

Diseases of the heart and the circulatory system, so-called cardiovascular diseases (CVD), are the main cause of death in Europe: accounting for over 4 million deaths each year [13]. The main cause of CVD is a disease of arteries, called atherosclerosis, in which plaque (a fatty substance) is deposited on the inside of the artery walls. Depositing plaque gradually causes narrowing the arteries. This narrowing prevents the blood from flowing properly through the arteries.

Coronary arteries are special blood vessels that supply the heart with necessary oxygen and nutrients. If they are being narrowed, the heart does not function properly without enough oxygen and nutrients. The result is coronary heart disease (CHD). CHD belongs to the most frequent forms of CVD [13]. Despite the fact that an essential decrease was registered in CHD mortality in a majority of countries (Figure 1), CHD belongs to the main causes of death. In the Czech Republic, age-directly standardized mortality (computed per 100 000 European standard population) from CHD decreased from 543 in 1986 to 328 in 1998 in men aged 35–74 years and from 202 to 120 in women.

Risk factors of CHD are summarized in Table 1 [14]. A successful cardiovascular preventive programme should result in a decrease of cardiovascular incidence, and consequently, mortality. The two aims should be targeted: to popularize health life style, and to identify persons with cardiovascular risk factors already present and to intervene their risk factors. The modifiable risk factors can be controlled by changing lifestyle or by pharmacotherapy.

Epidemiologists, statisticians and other health workers have been working on statistical models which produce an absolute risk estimation of developing CVD. The absolute risk is the probability of developing



<sup>1</sup> Standardized using the European standard population  
<sup>2</sup> CHD coded as 410–414 of ICD-8 and ICD-9, as I20–I25 of ICD-10, ICD International Classification of Diseases (8th, 9th and 10th Revisions)  
<sup>3</sup> For the Czech Republic 1986 instead of 1981  
 Data source: [13]

Figure 1: Age-directly standardized<sup>1</sup> mortality from coronary heart disease<sup>2</sup> (CHD), for adults aged 35-74 years, 1981<sup>3</sup>, 1998

CVD event within a given time period. These statistical models are increasingly used to identify a population at high risk.

Table 1: Risk factors of coronary heart disease (CHD)

Major risk factors	Cigarette smoking
	Elevated blood pressure
	Elevated serum total (and LDL <sup>1</sup> ) cholesterol
	Low serum HDL <sup>2</sup> cholesterol
	Diabetes mellitus
Predisposing risk factors	Advancing age
	Obesity <sup>3</sup>
	Abdominal obesity <sup>4</sup>
	Physical inactivity
	Family history of premature CHD
Conditional risk factors	Ethnic characteristic
	Psychosocial factors
	Elevated serum triglycerides
	Small LDL particles
	Elevated serum homocysteine
	Elevated serum lipoprotein(a)
	Prothrombotic factors (e.g. fibrinogen)
	Inflammatory markers (e.g. C-reactive protein)

<sup>1</sup> Low-density lipoprotein

<sup>2</sup> High-density lipoprotein

<sup>3</sup> Obesity defined as Body Mass Index (weight[kg]/height[m]<sup>2</sup>) > 30.0 kg/m<sup>2</sup>

<sup>4</sup> Abdominal obesity defined as waist circumference >102 cm for men, and >88 cm for women

Source: [14]

Well-known statistical models are those derived in the Framingham heart study (FHS) [1], [2], [9], [30]. While this study is based on a population of the United States, other studies estimate the absolute CHD risk for European populations [3], [7], [23], [28],

Although statistical models are developed for specific sub-populations, there are also applied in other populations. Generalization of the statistical models to external populations should be done cautiously. Epidemiological studies evaluate the validity of CVD risk estimations in other populations than that they were derived from.

## 2 Aims of the thesis

The aim of this work is

- to analyse cardiovascular risk factors in the Czech Republic, and above all
- to validate the following risk functions based on the Framingham and European populations in the Czech Republic (Table 2):
  - the Framingham CHD risk function (1991),
  - the Framingham CHD risk function (1998),
  - the SCORE fatal CVD risk function (2003).

The analysis of the risk factors and the validation of the risk functions were performed within the longitudinal primary prevention study of atherosclerotic risk factors, which is described in Materials and methods.



Table 2: Risk estimations based on the Framingham heart study (FHS) and the SCORE study

Study (year) [citation]	FHS (1991) [1]	FHS (1998) [30]	SCORE (2003) [7]
POPULATION	General USA (Framingham)	General USA (Framingham)	General, employees 12 European studies
Baseline examination	1968–1975	1971–1974	1967–1991
Gender (sample size)	Men(2 590) Women(2 983)	Men(2 489) Women(2 856)	Men (117 098) Women (88 080)
Age [yrs]	30–74	30–74	45–64
RISK FUNCTION			
Failure of interest <sup>1</sup>	CHD	CHD	Fatal CVD
Time until failure [yrs]	4–12	10	10
Statistical method	Weibull non-proportional hazards regression	Cox proportional hazards regression	Weibull proportional hazards regression
<i>Explanatory variables</i> <sup>2</sup>			
Gender	+	+	+
Age	+	+	+
Blood pressure (BP)		+	
Systolic BP	+		+
Cigarette smoking	+	+	+
Total cholesterol	+	+	+
HDL-cholesterol	+	+	
Diabetes mellitus	+	+	
Left ventricular hypertrophy	+		

<sup>1</sup> *Coronary heart disease (CHD)* involves angina pectoris, coronary insufficiency, myocardial infarction, and fatal CHD, *Fatal cardiovascular disease (CVD)* involves cardiovascular death defined as ICD-9 codes 401–414, 426–443 (with the exception of the following ICD9-codes for definitely non-atherosclerotic causes of death: 426.7, 429.0, 430.0, 432.1, 437.3, 437.4 and, 437.5), 798.1 (instantaneous death) and 798.2 (death within 24h of symptom onset)

<sup>2</sup> The explanatory variables marked with + were used for modelling the risk

## 3 Material and methods

### 3.1 Design of the STULONG study

The longitudinal primary prevention study of atherosclerotic risk factors, so-called the STULONG study, was conducted by 2nd Dep. of Internal Medicine, 1st Faculty of Medicine and General Faculty Hospital, Charles University in Prague 2 in 1975–1999.

In 1975 total 2370 men aged 38–49 living in the 2nd district in the centre of Prague (Prague 2) were randomly selected from the electoral register. It was the 50 % sample of men of that age who were living in Prague 2 in 1975. Of 2370 invited men, 1417 (59.8 %) men answered the invitation and underwent entry examination in 1975–1979. *The entry questionnaire* included questions on demographic and personal data (marital status, education, working physical activity, leisure physical activity, smoking, alcohol drinking, coffee drinking, tee drinking, personal and family history, chest pain, lower limbs pain, breathlessness) and results of physical (height, weight, systolic blood pressure (BP), diastolic BP, skinfolds), laboratory (cholesterol level, triglyceridy, uric acid) and ECG (electrocardiography) measurements.

**Definition of the groups:** Each man was classified into one of three groups (normal, risk, pathological) according to health status and occurrence of the risk factors of atherosclerosis at the entry into the study; the definition of risk factors is corresponding to the period of the beginning of the study (Table 3). The design of the STULONG study is pictured in Figure 2 [5], [25].

*Normal Group* (NG) included men without any risk factors of atherosclerosis mentioned in Table 3, without CVD, without diabetes mellitus, without other serious disease not enabling long term follow-up and without pathological finding on ECG curve at the entry into the study. NG

Table 3: Risk factors of atherosclerosis at the entry into STULONG in 1975–1979

Positive family history	Mother or father died from cardiac infarction, stroke or suddenly (excluding accident) at the age $\leq 65$ years
Obesity	Brocca index (BI) $\geq 115$ %, where $BI = \text{weight}[\text{kg}] / (\text{height}[\text{m}] - 100) \cdot 100$ %
Cigarette smoking	Number of cigarettes a day $\geq 15$ , or ex-smoker less than one year who smoked $\geq 15$ cigarettes a day before
Hypertension	Blood pressure $\geq 160$ and/or 95 mm Hg in 2 of 3 measurements (2 measurements at the entry into the study, 3rd measurement within 190 days from the entry), or hypertension in anamnesis
Hypercholesterolaemia	Total cholesterol $\geq 260$ mg/100ml (6.7 mmol/l)

was randomly divided into two groups: normal group regularly examined (NGE,  $n = 40$ ) and normal group unexamined regularly (NGN,  $n = 236$ ).

*Risk group* (RG) included men with at least one of the risk factors of atherosclerosis (Table 3), without CVD, without diabetes mellitus and without other serious disease not enabling long term follow-up and without pathological finding on ECG curve at the entry into the study. RG was randomised into two subgroups: risk intervention group (RIG,  $n = 427$ ) and risk control group (RCG,  $n = 432$ ). Pharmacological and non-pharmacological intervention of the risk factors of atherosclerosis in RIG was performed by specialists from 2nd Dep. of Internal Medicine, RCG was under health care of general practitioners. Since early eighties of the last century, the groups RIG and RCG had been merging, mainly for ethic reasons [25].

*Pathological group* (PG) included men with CVD, with diabetes mel-

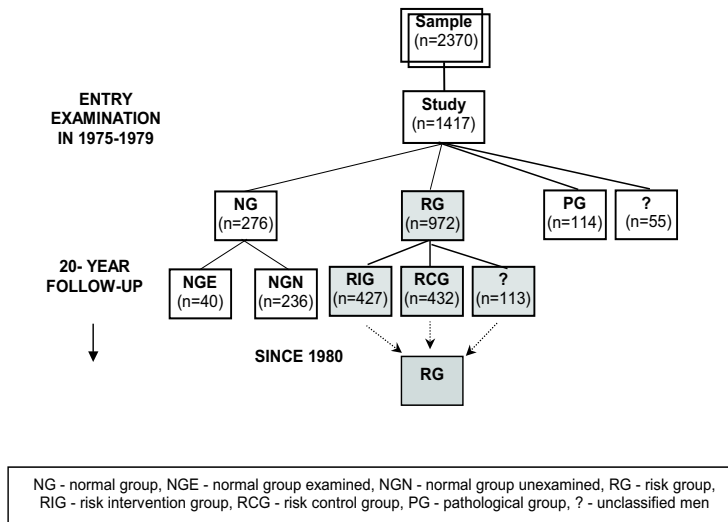


Figure 2: Design of the longitudinal primary prevention study of atherosclerotic risk factors (STULONG)

litus or with other serious disease not enabling long term follow-up or with pathological finding on ECG curve at the entry into the study.

**Evaluation of risk factors:** During the follow-up, control questionnaires were fulfilled by specialists from 2nd Dep. of Internal Medicine. *The control questionnaire* was consistent with the entry questionnaire except a question on family history excluded in the control questionnaire, and a question on feeding habits extra included in the control questionnaire.

**End-points:** During the 20-year follow-up period the first occurrence of atherosclerotic CVD was recorded. Atherosclerotic CVD were coded as D410–D414, D427, D430–D438, D440–D444 according to ICD-

8 (1975–1978), 410–414, 426–428, 431–438, 440–443 according to ICD–9 (1979–1993), and I20–I25, I44–I50, I60–I67, I69, I70 according to ICD–10 (1994–1999). CHD included ICD–8 codes D410–D414, ICD– codes 410–414, and ICD–10 codes I20–I25.

In 1999–2001, information on survival of men withdrawing or dropping out the study was ascertained (sources: outpatient departments, postal questionnaire). For deceased, date and cause of death were identified (sources: registry offices, Institute of Health Information and Statistics of the Czech Republic, relatives).

### **3.2 Validation studies within the STULONG study**

Each man of the STULONG study who met the following criteria was included in the validation study of the Framingham risk functions (1991, 1998) [1], [30]: classified into NG or RG at the entry into the study, and free of CHD at a control examination, and having information about the variables (Table 2) needed for the estimation risk. If more such control examinations were available, the first control examination, which met these conditions, was taken as a baseline.

We calculated individuals' 10-year absolute risk of CHD according to the Framingham risk function (1991 and 1998, respectively) on basis of their control examinations instead of their entry examination, because the level of HDL-cholesterol wasn't measured at the entry. The Framingham risk model (1991) requires information on the occurrence of left ventricular hypertrophy to estimate the 10-year absolute CHD risk. However, left ventricular hypertrophy was not surveyed at control examinations. Due to the fact that none of men from NG and RG suffered from left ventricular hypertrophy at the entry examinations, the risk was estimated on the assumption that left ventricular hypertrophy was not present at the control examinations.

Each man of the STULONG study who met the following criteria was included in the validation study of the SCORE risk function (2003) [7]: classified into NG or RG at the entry into the study, and having all information about the variables (Table 2) needed for the estimation of the risk of fatal CVD. Note that HDL-cholesterol is not required for the 10-year absolute fatal CVD risk estimation by this function. The SCORE risk function was validated on basis of entry questionnaires.

### 3.3 Statistical methods

A Web-based calculator for ROC curves [12], STATISTICA (StatSoft 1995), Egret (Cytel Software Corporation 1999) and R (Development Core Team 2003) software were used for statistical analysis of data. An attained level of significance  $p < 0.05$  was considered as a statistical significant;  $p$ -value for a two-tailed hypothesis test.

#### Association methods:

The Cox proportional hazards regression was used to determine baseline explanatory variables related to survival free of fatal atherosclerotic CVD. An important assumption of the Cox model is that hazards are proportional. This assumption was graphically assessed (the plot of  $\log(\text{time})$  versus the scaled Schoenfeld residuals), and statistically tested (the Pearson and Spearman correlation coefficients between  $\log(\text{time})$  and the scaled Schoenfeld residuals, and the test of significance of the interaction between  $\log(\text{time})$  and each explanatory variables).

The Kaplan-Meier method was used to estimate the survival free of fatal atherosclerotic CVD. The log rank test was applied to compare survival functions of groups defined according to the number of present risk factors at the entry. The analysis of variance (ANOVA) test was

used to compare mean age in the groups defined according to the number of present risk factors at the baseline.

### **Validation methods:**

Overall goodness of fit evaluated by tests of calibration and discrimination measures the degree of the accuracy of the prediction of a model.

*Calibration* of a model describes how closely the estimated risk agree with that observed. Calibration of the Framingham and SCORE models was measured with the Hosmer-Lemeshow (H-S) goodness of fit test [17], and a version of H-S goodness of fit test for a survival data [10]. The Framingham risk function (1998) was recalibrated by the methods described by D'Agostino et al. [9]. Note that recalibration does not influence discrimination.

*Discrimination* of a model expresses the ability of the model to distinguish observations with a positive and a negative outcome. In STULONG, discrimination was evaluated by a co-called Receiver Operating Characteristics (ROC) curve. The area under the ROC curve was estimated by the Wilcoxon statistic  $W$  [15]. The web-based calculator for ROC curves [11] was used to generate ROC curves with 95% confidence intervals. Besides the Wilcoxon statistic, the area under the ROC curve was also estimated by a method for survival data [22].

When validating the Framingham risk function (1998), the regression coefficients of this function were also estimated using the data from the STULONG study. Afterward *the estimations of the regression coefficients were compared* between the Framingham and STULONG models by the test statistic  $z$  [9]. Homogeneity of risk factors' distribution between the STULONG and FHS studies was tested by the Pearson chi-square test, and the test on standardized residuals. The Student t-test was used to compare mean age between the STULONG and FHS studies.

## 4 Results

### 4.1 Cardiovascular risk factors in the Czech Republic

Prevalence of the atherosclerosis risk factors (defined in Table 3) at the entry into the study in 1975–1979 is pictured in Figure 3.

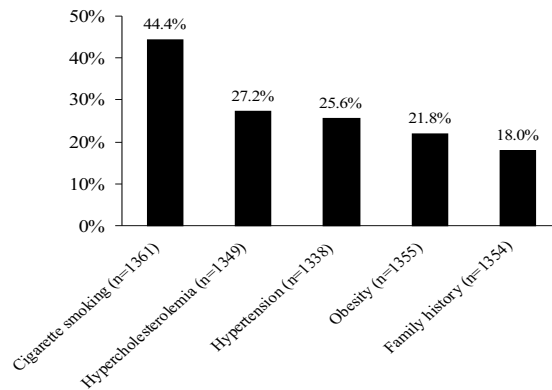


Figure 3: Prevalence of atherosclerosis risk factors at the entry into the study, 1975–1979

There was a significant ( $p < 0.001$ ) decrease in the survival free of fatal atherosclerotic CVD with the increasing number of the risk factors at the entry study in the risk group, Figure 4. Interestingly the groups did not differ ( $p = 0.759$ ) in mean age (46.2 years in all groups).

As written in Table 4, heavy smokers (at least 15 cigarettes daily) had hazard of fatal atherosclerotic CVD significantly higher than men with lower cigarette consumption, men with hypertension (blood pressure  $\geq 160/90$  mm Hg) had hazard significantly higher than men with blood pressure  $< 140/90$  mm Hg. Men with hypercholesterolemia (to-



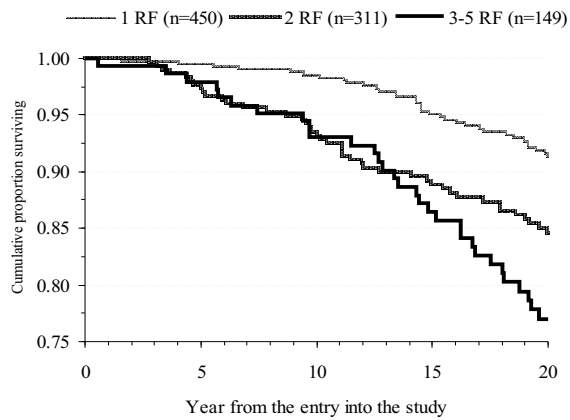


Figure 4: Kaplan-Meier survival free of fatal atherosclerotic CVD according to the number of risk factors (RF) in the risk group at the entry in to the study

tal cholesterol  $\geq 260$  mg/dl, i.e.  $\geq 6.7$  mmol/l) had hazard significantly higher than men with the level of cholesterol up to 202 mg/dl (i.e. 5.2 mmol/l). Men with university education had hazard significantly lower than men with elementary education.

## 4.2 Validation of cardiovascular risk estimations in the Czech Republic

### 4.2.1 Framingham risk function (1991)

Among 1 417 men aged 38–53 years at the entry, 1 248 men were from NG and RG. From these 1 248 men, 916 had the complete 10-year follow-up (CHD event within 10 years or the control examination at 10-year follow-up, or later), and the information on the risk factors, excluding HDL-cholesterol, needed for the estimation of an individual’s absolute

Table 4: Results of the Cox regression model: the numbers of men from RG in each category (n), hazard ratio (HR) of death from atherosclerotic CVD with 95% confidence interval (CI)

Variable	n	HR	95% CI	p-value	
Age [yrs]	926	1.1	1.1–1.2	<0.001	
Education	Elementary	110	1.0		
	Apprenticeship	274	0.7	0.4–1.3	0.273
	Secondary	292	0.6	0.3–1.0	0.050
	University	250	0.3	0.2–0.6	<0.001
Cigarette smoking	<15 cigarettes daily	396	1.0		
	≥15 cigarettes daily	530	3.0	2.0–4.6	<0.001
Blood pressure <sup>1</sup>	<140/90 [mm Hg]	525	1.0		
	140/90–159/94 [mm Hg]	211	1.5	0.9–2.3	0.125
	≥160/95[mm Hg]	190	2.8	1.8–4.3	<0.001
T-cholesterol	<202 [mg/dl]	199	1.0		
	202–259 [mg/dl]	412	1.4	0.8–2.5	0.182
	≥260[mg/dl]	315	1.8	1.0–3.1	0.043

<sup>1</sup> If only one value (systolic or diastolic blood pressure (BP)) exceeds the limit, patient belongs into the higher category (e.g. patient with BP 130/92 mm Hg belongs into the category 140/90–159/94 mm Hg), BP: mean of two measurements

10-year CHD risk by the Framingham risk function (1991), Table 5. 387 (42.2 %) of 916 men were nonsmokers (according to the definition in Table 3).

In 1979–1988, 540 of 916 men were without evidence of CHD having the complete follow-up of 10-year and all information on the risk factors, excluding left ventricular hypertrophy, needed for the estimation of the risk. All 540 men were without diabetes mellitus, and 271 (50.2 %) were actual nonsmokers (at least one cigarette a day). Statistical characteristics of men at the control examination (the baseline) are presented in Table 5.

When estimating the risk within 10 years from the baseline, the risk

Table 5: Baseline risk factors, FHS (1991)

Risk factors	n	Mean	Std.Dev.	Median	Min	Max
At the entry into the study <sup>1</sup>						
Age [yrs]	916	46.1	3.6	46.5	38.0	53.0
Systolic BP [mm Hg]	916	130.4	17.3	127.5	80.0	210.0
Diastolic BP [mm Hg]	916	83.4	11.2	82.5	50.0	132.5
T-cholesterol [mg/dl]	916	234.9	45.8	231.0	112.0	470.0
At the control examination						
Age [yrs]	540	51.1	3.6	51.0	44.0	62.0
Systolic BP [mm Hg]	540	133.0	16.9	130.0	90.0	200.0
Diastolic BP [mm Hg]	540	85.6	10.1	85.0	60.0	115.0
T-cholesterol [mg/dl]	540	224.6	40.2	222.0	128.0	391.0
HDL-cholesterol [mg/dl]	540	53.9	14.2	52.0	16.0	114.0

<sup>1</sup> BP - blood pressure: mean of two measurements, HDL-cholesterol not measured

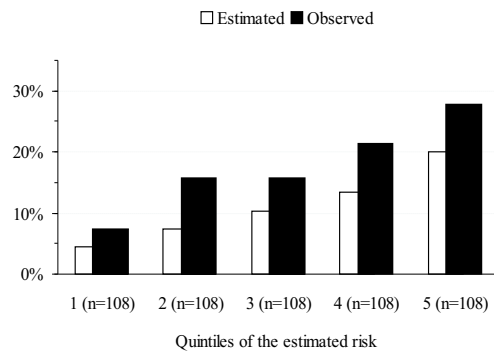


Figure 5: Observed and estimated risks of coronary heart disease (CHD) within 10 years, FHS (1991)

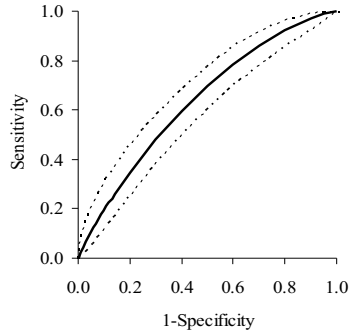


Figure 6: Receiver operating curve with 95% confidence interval (n=540), FHS (1991)

of CHD was estimated under the assumption that left ventricular hypertrophy was not present. At 10-year follow-up the estimated number of CHD events (60.1) was lower than observed (95). The Framingham risk function significantly underestimated the CHD risk observed across all quintiles of the risk ( $p < 0.001$ ), Figure 5. The trend in the proportion of CHD was significantly increasing across quintiles, ( $p < 0.001$ ).

Figure 6 shows the true positive fraction (sensitivity) of the Framingham risk function versus the false positive fraction (1-specificity) with the 95% confidence interval of the fitted ROC curve. According to the ROC analysis the Framingham risk function classified men at the baseline into those with and without developing CHD in the 10-year period with the accuracy of 62.8 % (95% CI 56.3 %, 69.3 %).

#### 4.2.2 Framingham risk function (1998)

A total of 646 men underwent the control (baseline) examination in 1979–1988.

Table 6: Baseline risk factors in FHS compared with STULONG, FHS (1998)

Risk factors	FHS (n=2489)	STULONG (n=646)	p-value <sup>1</sup>
Mean age (SD)	48.6 (11.7)	51.2 (3.7)	0.999
Blood pressure <sup>2</sup> [mm Hg]			
Optimal	<b>20.0</b> %	<b>11.3</b> %	
Normal	<b>24.0</b> %	<b>17.8</b> %	
High normal	20.0 %	16.1 %	<0.001
Hypertension stage I	<b>22.8</b> %	<b>39.0</b> %	
Hypertension stage II–IV	13.1 %	15.8 %	
Cigarette smoking [No/Yes]			
No	59.5 %	48.9 %	
Yes	40.5 %	51.1 %	<0.001
Diabetes mellitus [No/Yes]			
No	94.8 %	100.0 %	
Yes	5.2 %	0.0 %	<0.001
T-cholesterol [mg/dl]			
<160	<b>7.4</b> %	<b>4.2</b> %	
160–199	<b>31.2</b> %	<b>21.1</b> %	
200–239	38.9 %	41.2 %	<0.001
240–279	<b>16.7</b> %	<b>25.7</b> %	
≥280	5.8 %	7.9 %	
HDL-cholesterol [mg/dl]			
<35	<b>19.3</b> %	<b>6.5</b> %	
35–44	<b>35.5</b> %	<b>20.1</b> %	
45–49	14.9 %	16.3 %	<0.001
50–59	<b>19.6</b> %	<b>26.9</b> %	
≥60	<b>10.7</b> %	<b>30.2</b> %	

<sup>1</sup> For risk factors with more than two categories, bold font indicates the categories significantly contributed to overall significance

<sup>2</sup> For classification of blood pressure see [30]

They were significantly different in the background risk factors from those from Framingham. In STULONG, there was a higher prevalence of hypertensives, smokers, and men with hypercholesterolemia and with a higher HDL-cholesterol level, Table 6. None of men in STULONG suffered from diabetes mellitus. Out of 646 men, 450 men were censored at

10-year follow-up, 99 were censored before the 10-year follow-up without CHD, and 97 men were diagnosed with CHD in 10-year follow-up.

Unlike FHS, total cholesterol and HDL-cholesterol levels were not significantly associated with CHD risk in STULONG, Table 7. Smokers had significantly higher hazard of CHD event than non-smokers. The difference in the hazard ratio of smokers to non-smokers between FHS and STULONG was significant, ( $p=0.036$ ).

For each man of 646, 10-year absolute risk of CHD was estimated according to the Framingham risk function (Table 7). The mean 10-year absolute risk of CHD was of 12.8 % ( $n=646$ ). A total of 646 men were categorized into five groups according to quintiles of the estimated risk. There was the significant difference between the observed and estimated risks of CHD across quintiles ( $p=0.013$ ), Figure 7. Overall 12.8 % CHD events were estimated and 16.4 % observed in 10-year follow-up. The occurrence of CHD events observed in 10-year follow-up was significantly increasing across quintiles of the estimated risk ( $p < 0.001$ ).

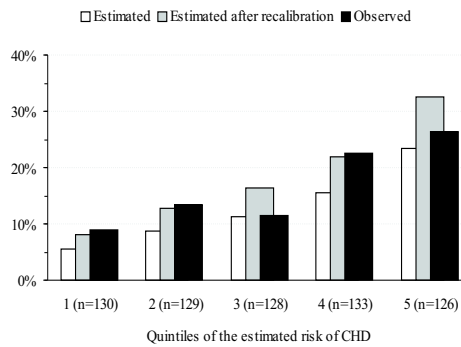


Figure 7: Observed and estimated risks of coronary heart disease (CHD) within 10 years, FHS (1998)

Table 7: Hazard rate (HR) of coronary heart disease (CHD) in FHS compared with STULONG, FHS (1998)

Risk factors	FHS (n=2 489)		STULONG (n=544)		<i>p</i> -value
	HR	95% CI	HR	95% CI	
Age [yrs]	1.05	1.04–1.06	1.05	0.99–1.11	0.999
Blood pressure <sup>1</sup> [mm Hg]					
Normal (including optimal)	1.00	Referent	1.00	Referent	
High normal	1.31	0.98–1.76	1.85	0.96–3.56	0.346
Hypertension stage I	1.67	1.28–2.18	1.72	0.97–3.03	0.927
Hypertension stage II–IV	1.84	1.37–2.49	3.36	1.77–6.37	0.094
Cigarette smoking [No/Yes]					
No	1.00	Referent	1.00	Referent	
Yes	1.68	1.37–2.06	2.84	1.82–4.41	0.036
Diabetes mellitus [No/Yes]					
No	1.00	Referent	1.00	Referent	
Yes	1.50	1.06–2.13	-	-	
T-cholesterol [mg/dl]					
<200	1.00	Referent	1.00	Referent	
200–239	1.31	1.01–1.68	0.92	0.54–1.57	0.243
≥240	1.90	1.47–2.47	1.35	0.79–2.28	0.260
HDL-cholesterol [mg/dl]					
<35	1.47	1.16–1.86	0.85	0.38–1.87	0.201
35–59	1.00	Referent	1.00	Referent	
≥60	0.56	0.37–0.83	0.72	0.45–1.15	0.432

<sup>1</sup> For classification of blood pressure see [30]

The 10-year survival free of CHD events was 90.0 % for FHS, and 83.6 % for STULONG. Survival rate, prevalence of risk factors and mean age in men from STULONG (Table 6) were used to recalibrate the Framingham risk function. The estimated risk of CHD by this recal-

ibrated function (not stated here because of limited space) was insignificantly different from that observed across quintiles of risk ( $p=0.320$ ), Figure 7.

Figure 8 shows the ROC curve with the 95% confidence interval of the fitted ROC curve ( $n=547$ ). Men who did not complete the 10 years of follow-up without having a CHD event were excluded from this analysis. The Framingham risk of CHD classified men free of CHD at the entry into those with and without CHD over 10 years with 63.2% accuracy, 95% CI (57.2 %, 69.3 %). When all men ( $n=646$ ) were included into the ROC analysis, the discrimination accuracy was of 63.8 %, 95% CI (58.4 %, 69.1 %).

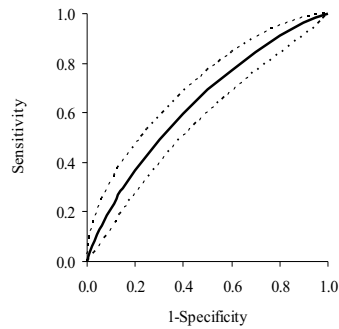


Figure 8: Receiver operating curve with 95% confidence interval ( $n=547$ ), FHS (1998)

#### 4.2.3 SCORE risk function (2003)

In the STULONG study, there were 1129 men free of CVD at the entry into the study, and having all information on variables needed to estimate the fatal CVD risk by the SCORE risk function, Table 8. 53.5 % of them



were actual smokers (at least one cigarette a day).

Table 8: Baseline risk factors, SCORE (2003)

Risk factors	n	Mean	Std.Dev.	Median	Min	Max
Age [yrs]	1129	46.1	3.6	46.0	38.0	53.0
Systolic BP <sup>1</sup> [mm Hg]	1129	130.5	17.6	130.0	80.0	220.0
Diastolic BP <sup>1</sup> [mm Hg]	1129	83.4	11.2	82.5	50.0	142.5
T-cholesterol [mg/dl]	1129	233.4	45.1	230.0	112.0	470.0

<sup>1</sup> BP - blood pressure: mean of two measurements

1129 men completed the 10 years follow-up (fatal CVD event within 10 years, or the control examination at 10-year follow-up or later). In the 10-year follow-up from the entry, the estimated number of fatal CVD was significantly differed from that observed across quintiles of the estimated risk ( $p=0.006$ ), Figure 9. Overall there was a total of 45 fatal CVD events observed and 28.0 estimated. The proportion of fatal CVD event observed in the 10-year follow-up was significantly increasing ( $p < 0.001$ ) across quintiles of the estimated risk.

Figure 10 shows the ROC curve with the 95% confidence interval. The fatal CVD risk estimated by the SCORE risk function classified men free of CVD at the entry into those with and without fatal CVD in the 10-year follow-up with 73.6% accuracy, 95% CI (66.4 %, 80.8 %).

When we added to 1129 men 81 men who were censored before the 10-year follow-up, calibration and discrimination accuracy of the SCORE risk function (n=1210) were similar as those for the complete data (n=1129): the risk function underestimated the observed risk of fatal CVD ( $p=0.006$ ), with discrimination accuracy of 74.3 %, 95% CI (67.0 %, 81.5 %).

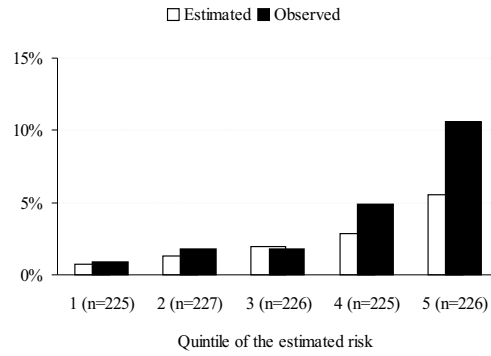


Figure 9: Observed and estimated risks of fatal cardiovascular diseases (CVD) within 10 years, SCORE (2003)

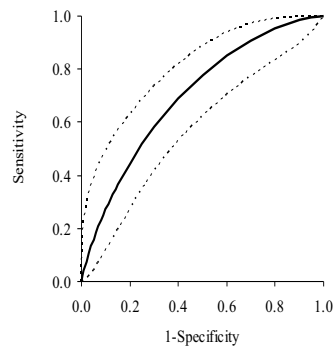


Figure 10: Receiver operating curve with 95% confidence interval (n=1129), SCORE (2003)

## 5 Discussion

Cardiovascular prediction models derived from a specific population may not hold for another population. It can happen if the two populations are not homogenous with respect to cardiovascular risk factors, and consequently, in the occurrence of cardiovascular events.

While the Framingham models are based on subjects from the town of Framingham (a suburb west of Boston, USA), the SCORE model on subjects from the European countries. Although these models were derived from specific populations, they are used to estimate cardiovascular risk in individuals from other populations. In this thesis we evaluate the accuracy of the Framingham CHD prediction models (1991, 1998) and the SCORE fatal CVD prediction model (2003) in Czech men from Prague.

The CHD mortality is higher in the Czech Republic than the USA and the majority of European countries, Figure 1. In STULONG, the mortality from CVD was significantly associated with known risk factors (hypertension, smoking, hypercholesterolemia, lower education), Table 4. There was detected a significant decrease in the survival free of fatal CHD event with the increasing number of the risk factors, Figure 4.

### 5.1 Validation studies in the Czech Republic

The Framingham (1991, 1998) and SCORE (2003) models underestimated the real absolute risk of the disease of the interest in STULONG. When recalibrating (i.e. adjusting) the Framingham CHD risk function (1998) for mean age, prevalence of risk factors and survival rate in the STULONG study, the observed and estimated risks of CHD were insignificantly different. Discrimination accuracy of the Framingham and SCORE models in STULONG was over 60 %.

The interactions between genes, lifestyle, and environmental factors may play an important part in the differences in the observed and estimated risks in the present validation study. Prevalence of the risk factors used for the estimation of the CHD risk by the Framingham risk function (1998) was not homogenous in the FHS and STULONG populations (Table 6). The hazard of CHD for smokers was essentially higher in STULONG than in FHS (Table 7).

The estimation of the risk by Framingham risk function (1991) may have been more precise, if the occurrence of left ventricular hypertrophy (LVH) was surveyed (LVH needed to estimate the CHD risk was assumed not to be present). In the case of the SCORE model, the definition of fatal CVD included somewhat different diseases than in the STULONG study (the definition of the diseases in Table 2 for SCORE, and on the page 12 for STULONG). Even if some discrepancy in the definitions, the definitions largely overlapped. Besides the difference in the endpoints, there were also some (in most cases only minor) differences in measurements of risk factors in the studies. It is only worth mentioning that persons who smoked regularly during the previous 12 months were classified as smokers in Framingham risk functions (1991, 1998), while in STULONG persons who smoked at least one cigarette a day.

STULONG was the primary preventive study. The discrepancy between the 10-year observed and estimated risks of CHD and fatal CVD, respectively, can indicate, among other things, that levels of the risk factors were non-randomly changing over the 10-year period. As it was shown by Boudik et al. [5], we can speculate about the efficiency of the intervention: by design a true control group was lacking.

In the STULONG study, diabetic men were excluded from the follow-up study according to the initial protocol. All subjects with diabetes identified during 20 years of follow-up were referred to outpatients for

diabetic department, but remain part of our survey. However, none diabetic man was recorded into the validation study. In 2002 diabetes afflicts 6.5 % of the Czech population, while in 1993 it was only 4.8 % [8].

The men from STULONG involved into the validation study do not represent all men from the Czech Republic. The STULONG study recruited middle-aged men from the centre of Prague, and a response rate was of 59.8 % in 1975–1979.

## 5.2 Other validation studies

The Framingham risk function (1991) [1] overestimated the risk of CHD for the Italian rural man population [19], and Western Europe [4]. While in the Czech population and aboriginal Australians [29], the Framingham risk function (1991) underestimated the absolute CHD risk. In England the Framingham risk function (1991) underestimated the 20-year absolute CHD risk for subjects with the lower absolute risk [24]. However, the Framingham function (1991) was derived to estimate the risk within 4–12 years, and here used for the 20-year period. The Framingham risk function (1991) overestimated the CHD and fatal CHD risk in British men [6], and the risk of fatal CHD and non-fatal myocardial infarction in Germany [16].

The Framingham risk function (1998) [30] underestimated the absolute risk of CHD in the low-risk group and overestimated in the high-risk group in healthy veterans in Boston (USA) [21], and overestimated the risk in men in Northern Ireland and France [12]. The Framingham risk model appropriately estimated the risk of CHD in Japanese man workers [27]. However, the Framingham risk function (1998) was developed to provide the 10-year CHD risk, and Japanese men were followed-up from 5 to 7 years.

The estimation of the hard CHD risk (hard CHD involves fatal CHD

and non-fatal myocardial infarction), by the Framingham risk function (2001) [9], within 5 years of follow-up for white and black men and women from the Atherosclerosis risk in communities study (a study in the United States) was reasonably good [9]. However, overestimation was observed in Japanese American and Hispanic men, Native American women [9], Chinese population [18], and Northeast Spain [20]. In the last mentioned study, the Framingham risk function (2001) was applied to the Gerona (Northeast Spain) population, but the prevalence of risk factors was estimated on the base of a cross-sectional study.

The discrimination ability, when calculated, was at least of 60 % in these studies. So that, even if calibration accuracy of the Framingham risk functions were not satisfying, the Framingham risk functions were able to rank individuals according to risk from low-risk to high-risk groups, with discrimination of 60 % and more.

Generally, there were a large geographic variation in coronary morbidity and mortality across the validation studies. Some of them used the Framingham functions to estimate CHD risk beyond the designed period and age range (e.g. the validation studies by Ramachandran et al. [24]). A great number of the validation studies restricted to a narrower age range than FHS. Some studies verified the risk estimation in samples recruited from structurally different populations than was the Framingham population. Remind that the FHS study recruited participants from residents of the town Framingham, however, some validation studies from e.g. rural populations and employees (e.g. the validation study by Menotti et al. [19]).

Note that we did not identify any study on external validation for the SCORE risk function by the reason that the model was issued only in 2003.

## 6 Conclusions

The Framingham (1991, 1998) and SCORE (2003) risk functions significantly underestimated an individual's 10-year absolute risk of CHD and fatal CVD, respectively, in middle and upper aged men (age range 44–62 years) from the Czech Republic (Prague). It seems that the underestimation was largely caused by differences in the background risk factors and frequency in the occurrence of disease outcome in populations under investigation. Recalibration of the Framingham risk function (1998) essentially increased the accuracy of the estimate of the CHD risk.

Despite these facts, the proportion of CHD and fatal CVD was significantly increasing across quintiles of the estimated risk for the Framingham and SCORE functions, respectively. The functions were able to rank individuals according to risk from low-risk to high-risk groups with the discrimination ability over 60 %. On the other hand the discrimination ability varying about 60 % can be debatable because it is not high.

The results have policy implications concerning the cardiovascular prediction models. Generally speaking, it is not surprising that the risk functions derived for a specific population will not be accurate in other populations. They can underestimate or overestimate the real absolute risk of a disease of the interest, if the populations are not homogenous with respect to risk factors (traditional, non-traditional) for cardiovascular diseases, and consequently, in the occurrence of cardiovascular diseases events. Then the risk functions must be recalibrated, or a new risk function derived. The risk functions can partly help in searching a population at high cardiovascular risk, if discrimination ability is sufficiently high.

## References

- [1] Anderson KM, Wilson PW, Odell PM, Kannel WB. An Updated Coronary Risk Profile, A Statement for Health Professionals. *Circulation*, 83:356–362, 1991.
- [2] Anderson KM, Odell PM, Wilson PW, Kannel WB. Cardiovascular Disease Risk Profile. *American Heart Journal*, 121:293–298, 1991.
- [3] Assmann G, Cullen P, Schulte H. Simple Scoring Scheme for Calculating the Risk of Acute Coronary Events Based on the 10-year Follow-up of the Prospective Cardiovascular Munster (PROCAM) Study. *Circulation*, 105:310–315, 2002.
- [4] Bastuji-Garin S, Deverly A, Moyses D et al. Intervention as a Goal in Hypertension Treatment Study Group. *Journal of Hypertension*, 20:1973–1980, 2002.
- [5] Boudík F, Reissigová J, Hrach K, Tomečková M, Bultas J, Anger Z, Aschermann M, Zvárová J: Primary prevention of coronary artery disease among middle aged men in Prague: twenty-year follow-up results, *Atherosclerosis* 184:86–93, 2006. (Translated into Czech: *Vnitřní lékařství* (in press), 2006.)
- [6] Brindle P, Emberson J, Lampe F, Walker M, Whincup P, Fahey T, Ebrahim S. Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *British Medical Journal*, 327:1267, 2003.
- [7] Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, De Bacquer D, Ducimetiere P, Jousilahti P, Keil U, Njolstad I, Oganov RG, Thomsen T, Tunstall-Pedoe H, Tverdal A, Wedel H, Whincup P, Wilhelmssen L, Graham IM; SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal*, 24:987–1003, 2003.
- [8] Czech Health Statistics Yearbook 2002. *Institute of Health Information and Statistics of the Czech Republic*, Prague 2003.
- [9] D’Agostino RB, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham Coronary Heart Disease Prediction Scores: Results of a Multiple Ethnic Groups Investigation. *Journal of the American Medical Association*, 286:180–187, 2001.
- [10] D’Agostino RB, Nam B-H. Evaluation of the Performance of Survival Analysis Models: Discrimination and Calibration Measures. In *Handbook of Statistics Volume 23: Advances in Survival Analysis*: Balakrishnan N and Rao CR editors, *Elsevier*, Amsterdam 2004.
- [11] Eng J. ROC analysis: web-based calculator for ROC curves. Version 1.0.2, March 2004; Web: <http://www.rad.jhmi.edu/roc>.
- [12] Empana JP, Ducimetiere P, Arveiler D, Ferrieres J, Evans A, Ruidavets JB, Haas B, Yarnell J, Bingham A, Amouyel P, Dallongeville J; PRIME Study Group. Are the Framingham and PROCAM Coronary Heart Disease Risk Functions Applicable to Different European Populations?. The PRIME Study. *European Heart Journal*, 24:1903–1911, 2003.
- [13] European Cardiovascular Disease Statistics. *British Heart Foundation*, London 2000.



- [14] Grundy SM , Pasternak R, Greenland P, Smith S, Fuster V. Assessment of Cardiovascular Risk by Use of Multiple-Risk-Factor Assessment Equations. A Statement of Healthcare Professionals From the American Heart Association and the American College of Cardiology. *Circulation*, 100:1481–1492, 1999.
- [15] Hanley JA, McNeil BJ. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) curve. *Radiology*, 143:29–36, 1982.
- [16] Hense HW, Schulte H, Lowel H, Assmann G, Keil U. Framingham Risk Function Overestimates Risk of Coronary Heart Disease in Men and Women from Germany—Results from the MONICA Augsburg and the PROCAM Cohorts. *European Heart Journal* 24:937–945, 2003.
- [17] Hosmer DW, Lemeshow S. Applied Logistic Regression. *John Wiley & Sons, Inc.*, New York 1989.
- [18] Liu J, Hong Y, D’Agostino RB Sr, Wu Z, Wang W, Sun J, Wilson PW, Kannel WB, Zhao D. Predictive Value for the Chinese Population of the Framingham CHD Risk Assessment Tool Compared with the Chinese Multi-Provincial Cohort Study. *Journal of the American Medical Association*, 291:2591-2599, 2004.
- [19] Menotti A, Puddu PE, Lanti M. Comparison of the Framingham Risk Function-based Coronary Chart with Risk Function from an Italian Population Study. *European Heart Journal*, 21:348–350, 2000.
- [20] Marrugat J, D’Agostino R, Sullivan L, Elosua R, Wilson P, Ordovas J, Solanas P, Cordon F, Ramos R, Sala J, Masia R, Kannel WB. An Adaptation of the Framingham Coronary Heart Disease Risk Function to European Mediterranean Areas. *Journal of Epidemiology and Community Health*, 57:634–638, 2003.
- [21] Orford JL, Sesso HD, Stedman M, Gagnon D, Vokonas P, Gaziano JM. A Comparison of the Framingham and European Society of Cardiology Coronary Heart Disease Risk Prediction Models in the Normative Aging Atudy. *American Heart Journal*, 144:95–100, 2002.
- [22] Pencina MJ, D’Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Statistics in Medicine*, 23:2109–2123, 2004.
- [23] Pocket Guide to Prevention of Coronary Heart Disease. International Task Force for Prevention of Coronary Heart Disease. *Brm Bruckmeier Verlag*, Grünwald 2003.
- [24] Ramachandran S, French J M, Vanderpump MP, Croft P, Neary RH. Using the Framingham Model to Predict Heart Disease in the United Kingdom: Retrospective Study. *British Medical Journal*, 320:676–677, 2000.
- [25] Reissigová J, Tomečková M. Intervention of the Risk Factors of Atherosclerosis and Cardiovascular Mortality. A 20-year Primary Prevention Study from a Statistician’s Point of View. *Cor et Vasa*, 45:249–255, 2003 (in Czech).
- [26] Reissigová J, Zvárová J. The Framingham Risk Function Underestimated Absolute Coronary Heart Disease Risk in Czech Men. *Methods of Information in Medicine* (to be published).

- [27] Suka M, Sugimori H, Yoshida K. Application of the Updated Framingham Risk Score to Japanese Men. *Hypertension Research*, 24:685–689, 2001.
- [28] Thomsen TF, Davidsen M, Jorgensen HIT, Jensen G, Borch-Johnsen K. A New Method for CHD Prediction and Prevention Based on Regional Risk Scores and Randomized Clinical Trials; PRECARD and the Copenhagen Risk Score. *Journal of Cardiovascular Risk*, 8:291–297, 2001.
- [29] Wang Z, Hoy WE. Is the Framingham Coronary Heart Disease Absolute Risk Function Applicable to Aboriginal People? *Medical Journal of Australia* 182:66–69, 2005.
- [30] Wilson PWF, D’Agostino RB, Levy D et al. Prediction of Coronary Heart Disease using Risk Factors Categories. *Circulation*, 97:1837–1847, 1998.

## Appendix: Publications and presentations of the author

There are *the publications* on which the author of the thesis participated:

1. Reissigová J, Tomečková M. Intervention of the Risk Factors of Atherosclerosis and Cardiovascular Mortality. A 20-year Primary Prevention Study from a Statistician's Point of View. *Cor et Vasa*, 45:249–255, 2003 (in Czech).
2. Reissigová J, Tomečková M. State of the Art Coronary Heart Disease Risk Estimations Based on the Framingham Heart Study *Central European Journal of Public Health*, 13:180–186, 2005.
3. Boudík F, Reissigová J, Hrach K, Tomečková M, Bultas J, Anger Z, Aschermann M, Zvárová J: Primary Prevention of Coronary Artery Disease Among Middle Aged Men in Prague: Twenty-Year Follow-up Results. *Atherosclerosis*, 184:86–93, 2006. (Translated into Czech: *Vnitřní lékařství* (in press), 2006.)
4. Reissigová J, Zvárová J. The Framingham Risk Function Underestimated Absolute Coronary Heart Disease Risk in Czech Men. *Methods of Information in Medicine* (to be published).

Note that the results were also presented in conferences and seminars:

1. Reissigová J. Validation of Coronary Heart Prediction. Proceedings of the IX. PhD. Conference, Institute of Computer Science Academy of Sciences of the Czech Republic (Paseky nad Jizerou, Czech Republic, September 25–26, 2003), MATFYZPRESS, ISBN 80-86732-16-9, 89-95, 2003.
2. Reissigová J, Tomečková M, Zvárová J. External Validation of the Framingham Risk Function in Men with Coronary Heart Disease from the Czech Republic. Proceedings of the International Joint Meeting EuroMISE 2004 (Prague, Czech Republic, April 12–15, 2004), EuroMISE, ISBN 80-903431-0-4, 29, 2004.
3. Reissigová J. Estimations of Cardiovascular Disease Risk - A survey of our Results from 2004. Proceedings of the IX. PhD. Conference, Institute of Computer Science Academy of Sciences of the Czech Republic (Paseky nad Jizerou, Czech Republic, September 29–October 1, 2004), MATFYZPRESS, ISBN 80-86732-30-4, 101–106, 2004.
4. Reissigová J. Does the Accuracy of the Estimation of Coronary Heart Diseases Risk by the Framingham Risk Function Depend on HDL-Cholesterol? Proceedings of the Seminar of Informal Technology in Health Care (Prague, Czech Republic, December 6, 2004), EuroMISE, spol. s r.o., ISBN 80-903431-1-2, 135–141, 2004.