
Introduction

Cardiovascular disease (CVD) has a large number of risk factors, of which sex, body mass index (BMI), smoking, diabetes and family history are the best known. However, other risk factors are involved, like total serum cholesterol, homocysteine and creatinine levels.

Analyzing the relation between these risk factors and CVD is straightforward for binary risk factors like smoking (yes/no), diabetes (yes/no) or family history (yes/no). The relation between continuous variables and CVD is much less known. It is the purpose of this research to clarify these less known relations.

Methods

We have used a dataset containing 2400 patients who were known with Familial Hypercholesterolemia (FH). These patients were followed from birth onward, until they were diagnosed with CVD, or until 31 December 2004 when they remained without CVD. For each of these patients their Event Free Survival (EFS) was recorded, where an event was defined as the diagnosis of CVD. If no such diagnosis was made by the end of the study, this time was also recorded in combination with the fact that they suffered no CVD. We used SPSS to do basic analysis on the data.

We have used the R program to analyze the data. Missing values were converted, “yes/no” values and “male/female” were converted into 1/0 and for several columns the data-type had to be changed from text to numeric. We used univariate and multivariate Cox regression to determine significant variables. Non-linear regression was analyzed with splines.

Results

The results of our data analysis can be found in Appendix A. A total number of 782 patients were diagnosed with CVD, and 487 of them were male. Smoking, family history and Serum Apolipoprotein A (LP(a)) appear strong risk factors predisposing for CVD.

For four risk factors we produced Kaplan-Meier graphs. These (Figures 1-4) can be found in Appendix B. Putting labels to the different lines in the graphs was difficult in R, but we assume that the least positive lines are always correlated to the presence of hypertension, alcohol usage, smoking and being of the male gender respectively.

The results of the Cox Regression analyses can be found in Appendices C and D. What clearly shows from these tables is that several factors which are a risk factor in the univariate analysis (like length) are less important in the multivariate analysis, because they are (probably) strongly related to another variable. From the multivariate analysis we identified six significant variables. Four of these variables (gender, alcohol usage, hypertension and family history) are binary, the other two (High Density Cholesterol and creatinine) are continuous.

We assumed that CVD linearly depends on these two continuous variables. Spline analysis proved our assumption to be correct, as the p-values for the linearity are significant. The results of this analysis are in Appendix E.

Conclusion

Apart from the obvious risk factors, we found six factors which have significant impact on developing cardiovascular disease. Two of these six, HDL and creatinine appear to have a linear relationship to CVD.

Discussion

The Cox regression method we used to determine the impact of the several factors assumes that these factors have a linear relationship upon the development of CVD. This might not be the case for the factors that we did not test.

Some of the other six significant variables that have been identified by the multivariate regression have a large amount (up to 20.5% for HDL) of missing values. Had these parameters been filled with all low values, the outcome of our research could have been different.

Appendix A – Baseline Characteristics

Baseline characteristics			
Variable	No CVD (n = 1618)	CVD (n = 782)	Missing (%)
Male gender – n (%)	693 (43,1)	487 (62,3)	
Length – mean (sd)	172,5 (9,3)	172,3 (9,2)	320 (13,3)
Weight – mean (sd)	74,2 (13,5)	76,7 (12,9)	250 (8,5)
BMI – mean (sd)	24,8 (3,6)	25,8 (3,3)	343 (14,3)
Pack years – mean (sd)	4,9 (10,7)	9,5 (15,8)	855 (35,6)
Systolic bloodpressure – mean (sd)	133,4 (18,1)	138,3 (21,6)	37 (1,5)
Diastolic bloodpressure – mean (sd)	81,2 (10,2)	83,4 (11,0)	37 (1,5)
Glucose – mean (sd)	5,0 (0,9)	5,3 (1,2)	118 (4,9)
HBA1c – mean (sd)	5,6 (1,1)	6,1 (1,4)	930 (38,8)
serum total cholesterol – mean (sd)	9,5 (1,9)	9,7 (2,2)	248 (10,3)
serum high density lipoprotein cholesterol – mean (sd)	1,2 (0,4)	1,1 (0,3)	440 (18,3)
serum triglycerides – mean (sd)	1,7 (0,9)	2,1 (1,2)	374 (15,6)
serum thyroid hormone – mean (sd)	1,9 (1,8)	1,8 (1,9)	427 (17,8)
serum apolipoprotein(a) LP(a) – mean (sd)	291,2 (356,6)	440,5 (514,5)	702 (29,3)
serum homocysteine – mean (sd)	11,7 (6,2)	13,9 (12,3)	1306 (54,4)
serum creatinine – mean (sd)	79,7 (14,5)	84,5 (16,9)	71 (3,0)
Alcohol – n (%)			491 (20,5)
1 [yes]	1.006 (42,0)	407 (17,0)	
2 [no]	329 (13,7)	167 (7,0)	
3 [unknown]	0 (0)	0 (0)	
Roken – n (%)			234 (9,8)
0 [no]	444 (18,5)	124 (5,2)	
1 [yes]	1.003 (41,8)	595 (24,8)	
Hypertension – n (%)			24(1,0)
0 [no]	1507 (62,8)	639 (26,6)	
1 [yes]	96 (4,0)	134 (5,6)	
Diabetes – n (%)			
0 [no]	1567 (98,0)	695 (88,9)	
1 [yes]	51 (2,0)	87 (11,1)	
Hypercholesterolemia family – n (%)			
0 [no]	1240 (76,7)	657 (84,0)	
1 [yes]	378 (23,3)	125 (16,0)	
CVD<60 years in female family members – n (%)			322 (13,4)
0 [no]	1209 (50,3)	491 (20,5)	
1 [yes]	242 (10,1)	136 (5,7)	
CVD<60 years in male family members – n (%)			306 (12,8)
0 [no]	736 (30,7)	304 (12,7)	
1 [yes]	724 (30,2)	330 (13,8)	

Appendix B – Kaplan-Meier Graphs for known risk factors

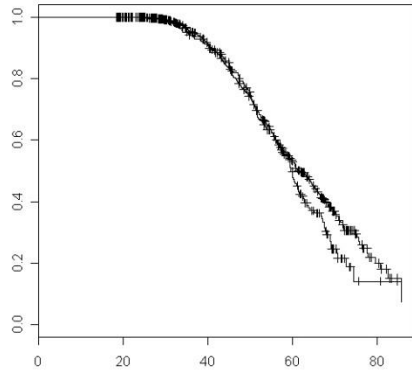


Fig 1. Survival - Hypertension

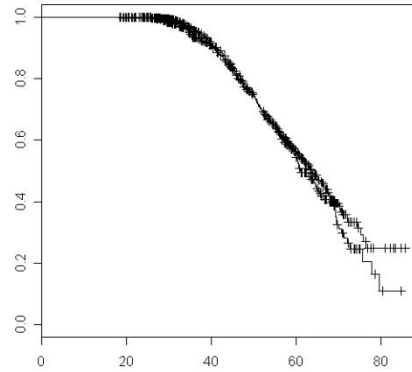


Fig 2. Survival – Alcohol usage

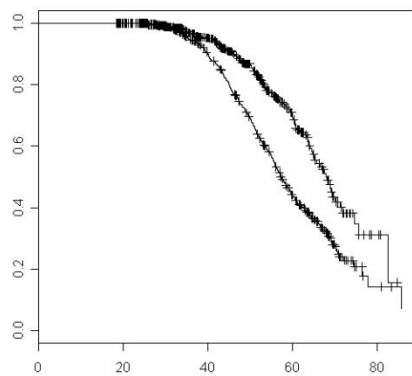


Fig 3. Survival – Smoking

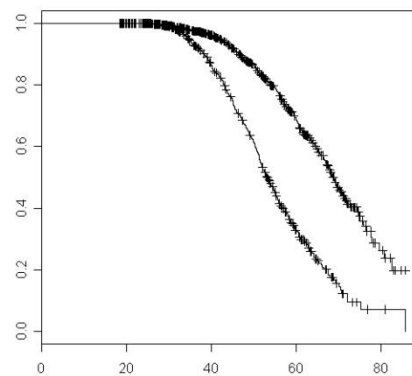


Fig 4. Survival – Gender

Appendix C – Univariate Cox Regression

	RR	SE(RR)	z	p	significance
geslacht	2.940	0.076300	14.2	0	*****
lengte	1.000	0.000114	1.11	2.7e-01	*
gewicht	1.020	0.002860	8	1.3e-15	*****
bmi	1.030	0.011300	2.6	9.3e-03	****
alcoholgebruik	1.110	0.092100	1.15	2.5e-01	*
roken	1.950	0.099200	6.74	1.6e-11	*****
pyears	1.010	0.002570	3.81	1.4e-04	*****
systbp	0.992	0.001820	-4.67	3.0e-06	*****
diasbp	0.992	0.003400	-2.35	1.9e-02	***
hypertension	1.160	0.095500	1.59	1.1e-01	*
Glucose	1.070	0.026900	2.55	1.1e-02	***
Hba1c	1.080	0.031400	2.33	2.0e-02	***
Tc	0.970	0.019300	-1.63	1.0e-01	*
HDL	0.349	0.138000	-7.63	2.4e-14	*****
Tg	1.120	0.033800	3.34	8.4e-04	*****
Tsh	1.000	0.000088	0.994	3.2e-01	*
Lpa	1.000	0.000168	2.41	1.6e-02	***
homocysteine	1.010	0.003050	4.78	1.8e-06	*****
creatinine	1.010	0.001900	4.52	6.2e-06	*****
diabetes	1.250	0.114000	1.93	5.3e-02	**
HCbijfamilie	1.110	0.097800	1.06	2.9e-01	*
anfavrev	0.752	0.097200	-2.93	3.4e-03	****
anfamaev	0.894	0.079700	-1.40	1.6e-01	*

Significance levels:

< 1e-03 *****

0.001 ****

0.01 ***

0.05 **

0.1 *

1

The variables in light blue have a significance of $p < 0.05$.

Appendix D – Multivariate Cox Regression

	RR	SE(RR)	z	p	significance
geslacht	5.910	0.375970	47.256	2.3e-06	*****
lengte	0.953	0.076149	-0.6339	5.3e-01	*
gewicht	1.070	0.082968	0.8179	4.1e-01	*
bmi	0.833	0.253374	-0.7215	4.7e-01	*
alcoholgebruik	2.377	0.262187	33.029	9.6e-04	*****
roken	1.030	0.286672	0.1018	9.2e-01	*
pyears	1.000	0.007568	0.0458	9.6e-01	*
systbp	0.992	0.007776	-10.010	3.2e-01	*
diasbp	0.989	0.012176	-0.8939	3.7e-01	*
hypertension	2.042	0.322684	22.119	2.7e-02	***
Glucose	0.944	0.180536	-0.3198	7.5e-01	*
Hba1c	1.166	0.136191	11.295	2.6e-01	*
Tc	0.925	0.063946	-12.110	2.3e-01	*
HDL	0.356	0.399684	-25.875	9.7e-03	****
Tg	0.897	0.118770	-0.9124	3.6e-01	*
Tsh	1.001	0.000557	10.580	2.9e-01	*
Lpa	1.000	0.000429	11.150	2.6e-01	*
homocysteine	1.035	0.019745	17.468	8.1e-02	**
creatinine	0.968	0.009847	-33.199	9.0e-04	*****
diabetes	0.461	0.694927	-11.144	2.7e-01	*
HCbijfamilie	1.943	0.332199	19.992	4.6e-02	***
anfavrev	0.702	0.278053	-12.726	2.0e-01	*
anfamaev	0.790	0.229985	-10.245	3.1e-01	*

Significance levels:

< 1e-03 *****

0.001 ****

0.01 ***

0.05 **

0.1 *

1

The variables in light blue have a significance of $p < 0.05$.

Appendix E – Spline Analysis

		coef	se(coef)	se2	Chisq	DF	p
pspline(HDL),	linear	-0.94252	0.13171	0.13156	51.21	1.00	8.3e-13
	nonlinear				7.09	3.01	7.0e-02
pspline(creatinine),	linear	0.00538	0.00253	0.00253	4.52	1.00	3.3e-02
	nonlinear				4.11	3.04	2.6e-01

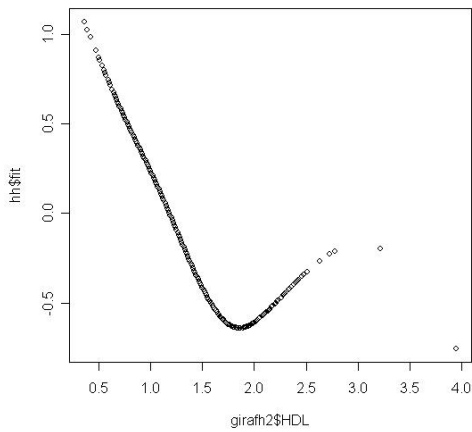


Fig. 5 Smoothing of High Density Cholesterol (HDL) with DF=4

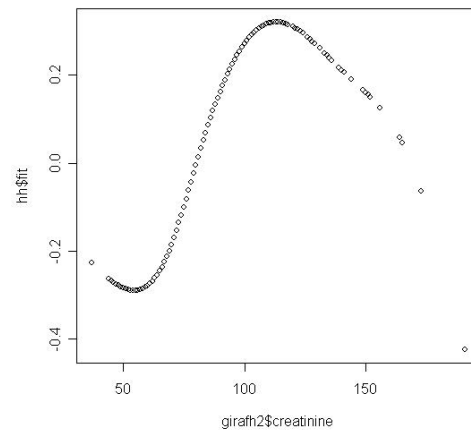


Fig. 6 Smoothing of Creatinine with DF=4