**eAppendix**

*Illustration of the models used by presenting annotated R code*

In this appendix, we provide detailed annotated R code of the statistical modeling, focusing on the estimation of natural direct, natural indirect, and total effects. We refer to the work of Pearl for definitions of these effects.1 We used methods proposed by VanderWeele et al. to obtain estimates of the effects.2–4

The data frame myData contains the following variables of individuals’ baseline examination. Data were extracted from the VHM&PP database.5 Individuals with a body mass index (BMI) <20kg/m2 were excluded.

BMI Body mass index in kg/m2

BMIclass Body mass index in 3 categories (1 – normal weight, 2 – overweight, 3 – obese)

sysBP Systolic blood pressure in mmHg

tChol Total cholesterol from a fasting blood test in mmol/l

lnGluc Logarithm of glucose from a fasting blood test in mmol/l

Age Age at baseline examination

Sex Sex (0 – female, 1 – male)

Smoking Smoking status in 3 categories (1 – non-smoker, 2 – former smoker, 3 – current smoker)

Time Event (death from coronary heart disease, defined via ICD-10 codes I20 to I25) time or censor time

CHDevent Event indicator (1 – yes, 0 – no)

In a first step, we assessed the interaction between the continuous variables BMI and Age on the outcome death from coronary heart disease using a Cox regression model (the command coxph is part of the survival package).

coxph(Surv(Time,CHDevent) ~ BMI\*Age + factor(Sex) + factor(Smoking), data=myData)

In a second step, we performed mediation analysis to obtain estimates of natural direct, natural indirect, and total effects, separately for the two age groups that are individuals <65 years and individuals ≥65 years at baseline examination, respectively.

We fit three separate linear regression models for each mediator conditional on BMI category and the three confounders age, sex, and smoking status. We decided to use logarithmized glucose values to improve model fit.

fit\_sysBP <- lm(sysBP ~ factor(BMIclass) + Age + factor(Sex) + factor(Smoking), data=myData, x=TRUE)

fit\_tChol <- lm(tChol ~ factor(BMIclass) + Age + factor(Sex) + factor(Smoking), data=myData, x=TRUE)

fit\_lnGluc <- lm(lnGluc ~ factor(BMIclass) + Age + factor(Sex) + factor(Smoking), data=myData, x=TRUE)

Then, a Cox regression model including the exposure BMI category, the three mediators, interactions with BMI category, and the three confounders was calculated.

fit\_cox <- coxph(Surv(Time,CHDevent) ~ factor(BMIclass) + sysBP + tChol + lnGluc + factor(BMIclass):sysBP + factor(BMIclass):tChol + factor(BMIclass):lnGluc + Age + factor(Sex) + factor(Smoking), data=myData)

The following commands extract coefficients and parameters of the four models needed for calculation of natural direct, natural indirect, and total effects.

alpha0 <- coef(fit\_sysBP)['(Intercept)']

alpha11 <- coef(fit\_sysBP)['factor(BMIclass)2']

alpha12 <- coef(fit\_sysBP)['factor(BMIclass)3']

alpha21 <- coef(fit\_sysBP)['Age']

alpha22 <- coef(fit\_sysBP)['factor(Sex)1']

alpha23 <- coef(fit\_sysBP)['factor(Smoking)2']

alpha24 <- coef(fit\_sysBP)['factor(Smoking)3']

delta1sq <- summary(fit\_sysBP)$sigma^2

beta0 <- coef(fit\_tChol)['(Intercept)']

beta11 <- coef(fit\_tChol)['factor(BMIclass)2']

beta12 <- coef(fit\_tChol)['factor(BMIclass)3']

beta21 <- coef(fit\_tChol)['Age']

beta22 <- coef(fit\_tChol)['factor(Sex)1']

beta23 <- coef(fit\_tChol)['factor(Smoking)2']

beta24 <- coef(fit\_tChol)['factor(Smoking)3']

delta2sq <- summary(fit\_tChol)$sigma^2

gamma0 <- coef(fit\_lnGluc)['(Intercept)']

gamma11 <- coef(fit\_lnGluc)['factor(BMIclass)2']

gamma12 <- coef(fit\_lnGluc)['factor(BMIclass)3']

gamma21 <- coef(fit\_lnGluc)['Age']

gamma22 <- coef(fit\_lnGluc)['factor(Sex)1']

gamma23 <- coef(fit\_lnGluc)['factor(Smoking)2']

gamma24 <- coef(fit\_lnGluc)['factor(Smoking)3']

delta3sq <- summary(fit\_lnGluc)$sigma^2

theta11 <- coef(fit\_cox)['factor(BMIclass)2']

theta12 <- coef(fit\_cox)['factor(BMIclass)3']

theta2 <- coef(fit\_cox)['sysBP']

theta3 <- coef(fit\_cox)['tChol']

theta4 <- coef(fit\_cox)['lnGluc']

theta51 <- coef(fit\_cox)['factor(BMIclass)2:sysBP']

theta52 <- coef(fit\_cox)['factor(BMIclass)3:sysBP']

theta61 <- coef(fit\_cox)['factor(BMIclass)2:tChol']

theta62 <- coef(fit\_cox)['factor(BMIclass)3:tChol']

theta71 <- coef(fit\_cox)['factor(BMIclass)2:lnGluc']

theta72 <- coef(fit\_cox)['factor(BMIclass)3:lnGluc']

We calculated the effects at the mean values of the confounders. The following lines of code give the mean values c1, c2, c3, and c4 of the respective variables.

c1 <- mean(fit\_sysBP$x[,'Age'])

c2 <- mean(fit\_sysBP$x[,'factor(Sex)1'])

c3 <- mean(fit\_sysBP$x[,'factor(Smoking)2'])

c4 <- mean(fit\_sysBP$x[,'factor(Smoking)3'])

Finally, the desired effects were obtained by combining these values as described by VanderWeele et al.2–4

Estimates for natural indirect effects (hazard ratio overweight vs. normal weight and obese vs. normal weight, respectively) were given as

HR\_NIE\_overweight <- exp(alpha11\*(theta2+theta51)+beta11\*(theta3+theta61)+gamma11\*(theta4+theta71))

HR\_NIE\_obese <- exp(alpha12\*(theta2+theta52)+beta12\*(theta3+theta62)+gamma12\*(theta4+theta72))

Estimates for natural direct effects (hazard ratio overweight vs. normal weight and obese vs. normal weight, respectively) were given as

HR\_NDE\_overweight <- exp(theta11+theta51\*(alpha0+alpha21\*c1+alpha22\*c2+alpha23\*c3+alpha24\*c4+theta2\*delta1sq)+theta61\*(beta0+beta21\*c1+beta22\*c2+beta23\*c3+beta24\*c4+theta3\*delta2sq)+theta71\*(gamma0+gamma21\*c1+gamma22\*c2+gamma23\*c3+gamma24\*c4+theta4\*delta3sq)+0.5\*(theta51\*\*2\*delta1sq+theta61\*\*2\*delta2sq+theta71\*\*2\*delta3sq))

HR\_NDE\_obese <- exp(theta12+theta52\*(alpha0+alpha21\*c1+alpha22\*c2+alpha23\*c3+alpha24\*c4+theta2\*delta1sq)+theta62\*(beta0+beta21\*c1+beta22\*c2+beta23\*c3+beta24\*c4+theta3\*delta2sq)+theta72\*(gamma0+gamma21\*c1+gamma22\*c2+gamma23\*c3+gamma24\*c4+theta4\*delta3sq)+0.5\*(theta52\*\*2\*delta1sq+theta62\*\*2\*delta2sq+theta72\*\*2\*delta3sq))

The total effects were obtained as the product of the hazard ratios of the natural direct and natural indirect effects.

HR\_TE\_overweight <- HR\_NIE\_overweight\*HR\_NDE\_overweight

HR\_TE\_obese <- HR\_NIE\_obese\*HR\_NDE\_obese

Proportions on a logarithmized scale were calculated as follows:

HR\_NDE\_overweight\_pct <- log(HR\_NDE\_overweight)/log(HR\_TE\_overweight)\*100

HR\_NIE\_overweight\_pct <- log(HR\_NIE\_overweight)/log(HR\_TE\_overweight)\*100

HR\_NDE\_obese\_pct <- log(HR\_NDE\_obese)/log(HR\_TE\_obese)\*100

HR\_NIE\_obese\_pct <- log(HR\_NIE\_obese)/log(HR\_TE\_obese)\*100

The uncertainty of natural direct, natural indirect, total effects, and the proportions was estimated with the bootstrap method using the R package boot. For this purpose, we encapsulated the commands given above in an R function BootstrappingFunction and then performed the bootstrapping with 5000 iterations and extracted the bootstrap percentile intervals with the following commands:

CIs -> boot(myData, BootstrappingFunction, R = 5000, stype = "i")

boot.ci(CIs, type="perc")

1. Pearl J. Direct and Indirect Effects. Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence: Morgan Kaufmann Publishers Inc., 2001:411-420.

2. VanderWeele TJ. Causal mediation analysis with survival data. *Epidemiology*. 2012;22(4):582-585.

3. Vanderweele TJ. Mediation analysis with multiple versions of the mediator. *Epidemiology*. 2012;23(3):454-463.

4. VanderWeele TJ. *Explanation in Causal Inference: Methods for Mediation and Interaction*. 1st ed. Oxford University Press; 2015.

5. Ulmer H. Long-term tracking of cardiovascular risk factors among men and women in a large population-based health system The Vorarlberg Health Monitoring & Promotion Programme. *Eur Heart J*. 2003;24(11):1004-1013.