**eAppendix**

**1. Estimation of measures of additive interaction**

Consider model (1) i.e.

The combined effect of the arbitrary increment and of the continuous determinants X and Y is:

The effect of an arbitrary increment of X, for a specific value of Y:

The effect of arbitrary increment of Y, for a specific value of X:

By assuming that odds ratios (ORs) from (1) approximate relative risks (RR), it is straightforward that

Knol and colleagues1 estimated Rothman's indexes in case of continuous risk factors; however, their formulae hold only for increases by 1 unit of the 2 risk factors from a background level of exposure of =0 and =0.

**2. Estimation of variances and confidence intervals for indexes of additive interaction**

The partial derivatives of RERI, AP and ln(S) (*f, g* and *ln(h)* respectively)with respect to evaluated at , are shown below:

Relative excess risk due to interaction (RERI):

, so

(S.1)

(S.2)

(S.3)

Proportion attributable to interaction (AP):

,so

(S.4)

(S.5)

(S.6)

Synergy index (S): After converting S to ln(S), we have

so,

(S.7)

(S.8)

(S.9)

*Estimation of variances and confidence intervals for measures of additive interaction between two continuous risk factors using logistic regression*

For estimating confidence intervals for each one of the Rothmans indexes, we calculated the necessary variance estimators, by applying the delta method. Let denote the 3 estimated coefficients from model (1). Rothman’s indexes estimated in (3), (5), (7) in the Research Letter can be seen as functions of and , with parameters the background level of exposure x0 on X and y0 on Y, and the arbitrary increment of X and of Y, i.e:

(S.10)

(S.11)

(S.12)

The variances of (S.10) – (S.12) can then be estimated as follows:

, so

Where:

is the partial derivative of *f* with respect to evaluated at j=1,2,3, that is:

and , k=2,3 and k>j.

Similarly for AP :

,

,

where

For S, since it is a ratio, the coverage properties are better if a confidence interval for ln(S) rather than for S is calculated2. Therefore, at first a confidence region for ln(S) was formed and then the limits of this confidence region were exponentiated in order to calculate a confidence interval for S. So, ln(S) and its variance can be written as:

so

, that is

where

So the 95% confidence intervals for RERI, AP and S can be calculated as follows:

For RERI: ,

for AP: ,

and for S:

**3. Illustrative dataset**

We used data from the Greek EPIC cohort. EPIC is a European longitudinal study designed to investigate the relationships between diet, lifestyle and environmental factors and the incidence of cancer and other chronic diseases3. In our paradigm, our sample consists of adult women, aged less than 45 years old of the Greek segment of the EPIC study4.

The binary outcome we considered was whether a subject was hypertensive or not at enrollment (cross sectional analysis). The information of the dichotomous outcome was collected using as criteria if the participant’s diastolic or systolic blood pressure was measured more than 90 mmHg or 140 mmHg respectively at baseline. We further considered BMI (in kg/m2) and level of adherence in Mediterranean Diet (MD)2 at recruitment as continuous and ordinal risk factors respectively. The variable related to (non-) adherence to MD (NMD), has 3 levels; in the 1st level, we considered those participants that their MD score is 6-9 (i.e. high adherence to MD), in the 2nd level, those with MD score between 2-5 (i.e. moderate adherence to MD) and in the 3rd level, those with MD score 0-1 (i.e. low adherence to MD). Moreover, we include in the analysis, age (in years), height (in cm) and education (3 levels; categorically modeled) as possible confounders. With respect to the variable related to education, we considered in the 1st level those with elementary education, in the 2nd level those with secondary education and in the 3rd level those who have, at least, a university degree. Participants with missing values in any of the above variables were excluded this analysis. Descriptive statistics of all variables included in the analysis are shown in table S1.

We hypothesized that increasing BMI and increasing levels of NMD would result in increasing prevalence of hypertension. In order to test our hypothesis, we ran a logistic regression model with hypertension as the dependent variable, while BMI (variable X) and increasing levels of NMD (variable Y) are considered the risk factors of interest. Age (variable Z1), education (Z2 is the dummy variable corresponding to the comparison of the 2nd level of education vs the 1st and Z3 is the dummy variable for the comparison of the 3rd level of education vs the 1st) and height (variable Z4) are considered as potential confounders. Let us consider the following model (S.13)

(S.13)

where p is the probability of the occurrence of hypertension and the estimated coefficients of the model. This is the model without the interaction between X and Y in order to illustrate that BMI and NMD are risk factors for the occurrence of hypertension. The results of the logistic regression model (S.13) are presented in table S2. Both BMI and NMD are positively associated with higher occurrence of hypertension although only BMI is deemed statistically significant. Subsequently we estimated the measures for additive interaction between BMI and NMD on the prevalence of hypertension whilst adjusting for the effects of the above potential confounders. In order to estimate Rothman's measures for additive interaction, we ran the following logistic regression model (S.14) which is (1) of the Research Letter when n=4, i.e.

(S.14)

The results of model (S.14) [or (1)] are presented in the Research Letter in Table 1.

**Script. Stata code for estimating additive interaction between any variables X and Y on binary outcome d.**

\*Programming code for additive interaction between variables X, Y on binary outcome d

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\* At the beginning, the user has to define the variables d, x, y and x\_y (i.e. the \*product term x\*y), and any potential confounder that should be included in the \*logistic regression model.

\* Here we use 3 such variables, Z1, Z2, Z3

global d " hyper "

global x " bmi "

global y " nmd "

global x\_y " bmi\_nmd "

global z1 " age "

global z2 " i.education "

global z3 " height "

\* In the Research Letter, for our illustrative example, the variables we used are:

\* 1) d=hypertension (yes/no) --> hyper

\* 2) X=Body Mass Index (BMI), calculated in kg/m2 --> bmi

\* 3) Y=Non adherence to Mediterrenean diet NMD (ordered in three levels) --> nmd

\* 4) X\_Y=Product term of BMI and NMD --> bmi\_nmd=bmi\*nmd

\* 5) Z1=age (continuous in years) --> age

\* 6) Z2=education (categorically in 3 levels) -> i.education

\* 7) Z3=height (continuous in cm)--> height

\* The user has to state also the parameters x0, y0, dx and dy.

\* In our example, in the first line of table 1, we have used x0=25, y0=1, dx=2, dy=1

global x0 "25"

global y0 "1"

global dx "2"

global dy "1"

logit $d $x $y $x\_y $z1 $z2 $z3

\*To estimate RERI:

nlcom(RERI:exp($dx\*\_b[$x]+$dy\*\_b[$y]+($dx\*$dy+$dy\*$x0+$dx\*$y0)\*\_b[$x\_y])-exp($dx\*\_b[$x]+$dx\*$y0\*\_b[$x\_y])-exp($dy\*\_b[$y]+$dy\*$x0\*\_b[$x\_y])+1 )

\*the output give the point estimate of RERI for x0, y0, dx, dy, its associated \*standard error using the delta method and the 95% CI

\*To estimate AP:

nlcom ( AP:1-exp(-\_b[$y]\*$dy-\_b[$x\_y]\*($dx\*$dy+$x0\*$dy))-exp(-\_b[$x]\*$dx-\_b[$x\_y]\*($dx\*$dy+$y0\*$dx))+exp(-\_b[$x]\*($dx)-\_b[$y]\*($dy)-\_b[$x\_y]\*($dx\*$dy+$y0\*$dx+$x0\*$dy)) )

\*the output give the point estimate of AP for x0, y0, dx, dy, its associated standard \*error using the delta method and the 95% CI

\* To estimate S:

\*Since S is a ratio, one has to estimate ln(S) (named as "ln\_syn\_index" in the nlcom \*command below) and the SE for ln(S) for x0, y0, dx, dy:

nlcom (ln\_syn\_index:ln(exp($dx\*\_b[$x]+$dy\*\_b[$y]+($dx\*$dy+$dy\*$x0+$dx\*$y0)\*\_b[$x\_y])-1)-ln(exp($dx\*\_b[$x]+$dx\*$y0\*\_b[$x\_y])+exp($dy\*\_b[$y]+$dy\*$x0\*\_b[$x\_y])-2) ), post

\*After obtaining \_b[ln\_syn\_index] and \_se[ln\_syn\_index] from the above command we \*estimate S and the associated 95% CI as follows:

\*\*\*Estimate S for x0, y0, dx, dy

scalar Syn\_index=exp(\_b[ln\_syn\_index])

\*\*\*Estimate 95% CI for S, for x0, y0, dx, dy

scalar Syn\_index\_low95 =exp(\_b[ln\_syn\_index]-invnormal(0.975)\*\_se[ln\_syn\_index])

scalar Syn\_index\_high95=exp(\_b[ln\_syn\_index]+invnormal(0.975)\*\_se[ln\_syn\_index])

\*\*\*\*And present in the output these estimates for S in one line

mat define Synergy\_index=(Syn\_index, Syn\_index\_low95, Syn\_index\_high95)

mat rown Synergy\_index= Syn\_index

mat coln Synergy\_index= S\_index S\_low95 S\_high95

mat list Synergy\_index

\*\*\*\*End of programming code\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**References**

1. Knol MJ, van der Tweel I, Grobbee DE, Numans ME, Geerlings MI. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *Int J Epidemiol* 2007;**36**:1111-8.
2. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology* 1992;**3**:452-6.
3. Riboli E and Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;**26** Suppl 1:S6-14.
4. Trichopoulou A, Costacou T, Bamia C, Trichopoulos D. Adherence to a Mediterranean diet and survival in a Greek population. *N Engl J Med* 2003;**348**:2599-608

|  |  |  |
| --- | --- | --- |
| **Continuous variables** | | |
|  | **Mean** | **SD** |
| Age (in years) | 37.99 | 4.61 |
| BMI (in kg/m2) | 26.62 | 5.00 |
| Height (in cm) | 160.02 | 6.03 |
|  | | |
| **Categorical variables** | | |
|  | | **N (%)** |
| Hypertension | |  |
| No | | 4336 (90.24%) |
| Yes | | 469 (9.76%) |
| Non adherence to Mediterranean Diet (NMD) | |  |
| 1-->Med. Diet score 6-9 | | 1133 (23.58%) |
| 2-->Med. Diet score 2-5 | | 3448 (71.76%) |
| 3-->Med. Diet score 0-1 | | 224 (4.66%) |
| Education level | |  |
| 1-->Elementary | | 1407 (29.28%) |
| 2-->Secondary | | 1917 (39.90%) |
| 3-->University degree or higher | | 1481 (30.82%) |

**Table S1**: Descriptive statistics of age, body mass index (BMI), height, hypertension, non adherence in Mediterranean Diet (NMD) and education in 4805 women aged less than 45 from the Greek-EPIC cohort

**Table S2**: Estimated coefficients from the logistic regression model (S.13) of the prevalence of hypertension in women aged less than 45 yrs old

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Logistic regression model (S.13)** | | | | |
|  | b | se | Odds Ratio (eb) | 95% Confidence Interval for Odds Ratio |
| **Body Mass Index** (in kg/m2) | 0.130 | 0.009 | 1.14 | (1.12 - 1.16) |
| **Non adherence to Mediterranean Diet** (3 levels - ordered) | 0.166 | 0.105 | 1.18 | (0.96 - 1.45) |
| **Age** (in years) | 0.068 | 0.013 | 1.07 | (1.04 - 1.10) |
| **Education**: (reference category - elementary school) |  |  |  |  |
| secondary | -0.541 | 0.118 | 0.58 | (0.46 - 0.73) |
| university | -0.897 | 0.147 | 0.41 | (0.31 - 0.54) |
| **Height** (in cm) | 0.021 | 0.009 | 1.02 | (1.0 - 1.04) |