# 1 OBJECTIVE

Environmental factors such as diet are likely important for the aetiology of many cancers. Fruit and vegetables have been suggested to explain 5-12 % of all cancers (1). The recent report by the World Cancer Research Fund(WCRF) concludes that there is a probably protective effect of fruits and vegetables and their constituents against cancers of the eosaphagus, pharynx, larynx, stomach, lung, colorectal, pancreas and prostate (2). So far only four cohorts have evaluated the association between fruits and vegetables and total cancer incidence. Three of them reported null associations(3-5), while a recent analysis within the Greeek EPIC Cohort Study reported that in women vegetables and fruits were inversely associated with incidence of cancer overall (6), while no effect was found in men.

The primary objective of this study is to investigate the association between fruits and vegetable intake and total cancer incidence in a cohort of Swedish women. We hypothesise that fruit and/or vegetable intake is inversely associated to the total cancer incidence.

# 2 NOTATION

# 3 PROPOSED AUTHORS IN ORDER

Löf M, Sandin S, Lagiou P, Adami H-O, Trichopoulos D, Weiderpass E

# 4 EXCLUSION CRITERIA

The following subsets of women will be excluded from the analysis:

1. Subjects coded as dead before cohort enrolment
2. Subjects coded as emigrated before cohort enrolment
3. Missing values on vital status information (alive, dead or emigrated) at follow-up
4. Subjects with recorded cancer before cohort enrolment
5. Subject with "extremely high or extremely low" total energy intake (<1% and >99 % percentiles) (variable n2)

# 5 DATA MANAGEMENT

## *5.1 Datasets*

In the analyses we will utilize the following datasets:

- nudes1_1_3: The MEB nutritonal data
- qitem: A nudes table with all unique dishes from the food frequency questionnaire
- can0612: Cancers (from the Swedish cancer register) up to 31-Dec-2006
- emim0612: Emigration and immigration data (from Swedish bureau of statistics) up to 31-Dec-2006
- dead0612: Date of death (from Swedish bureau of statistics) up to 31-Dec-2006

Nudes is the MEB database for management and calculation of food intake and nutrients.

### 5.2 DEFINITION OF TOTAL CANCER EVENT

For the present analysis a cancer event is defined by:

1. Data from the Swedish cancer registry complete up to 31 Dec 2006 will be utilised.
2. Non-melanoma skin cancers (ICD7=191 both benign and malign tumours) are not considered as a cancer event.
3. Cancer is here defined as a registered tumour coded as the ICD7 codes given in Appendix 1.
4. Given the prerequisites above (no 1-3) only the first cancer event will be considered.

## 6 PLANS FOR STATISTICAL ANALYSES

### 6.1 VARIABLES TO ANALYSE

Below are the variables to be considered in the analyses listed.

<u>Variables to obtain from the database(s)</u>

1. apple: apple (gram/day)
2. banana: banana (gram/day)
3. juice: juice (gram/day)
4. orange: orange (gram/day)
5. broccoli: broccoli (gram/day)
6. cabbage: cabbage (gram/day)
7. carrot: carrot (gram/day)
8. cauliflower: cauliflower (gram/day)
9. onion: onion (gram/day)
10. root: root (gram/day)
11. spinach: spinach (gram/day)
12. tomato: tomato (gram/day)
13. pea: pea (gram/day)
14. bean: bean (gram/day)
15. beef: beef (gram/day)
16. blacksausage: blacksausage (gram/day)
17. groundmeat: groundmeat (gram/day)
18. liver: liver(gram/day)
19. liverpaste: liverpaste (gram/day)
20. pork: pork (gram/day)

21. sandwichmeat: sandwichmeat (gram/day)
22. sausage: sausage (gram/day)
23. N2: Total energy intake (kJ/day)
24. N29: Alcohol=ethanol (gram/day)
25. N14: Alfa-tokoferol (mg/day)
26. N15: Vitamin C (mg/day)
27. N52: Folate (mg/day)
28. N51: Fibre (g/day)
29. ICD-7: ICD-codes from the Swedish Cancer registry
30. DIAG_DAT: Date of diagnosis associated with the ICD-7
31. BIRTH_DAT: Date of birth of the subject
32. Date for emigration
33. Date for immigration
34. Date of death
35. PARITY: Number of children. Categories 0, 1, 2, 3 ≥4 children
36. WEIGHT: Body weight (kg) at enrolment
37. HEIGHT: Height (cm) at enrolment
38. x-multivitamin: Use of dietary supplements. Categories: yes or no (This variable is in a dataset called vitamin_d created by Pouran)

Derived variables

1. BMI: BMI at enrolment
2. ATT_AGE: Attained age at follow-up. Age at emigration, death, cancer or 31st Dec 2006, whatever comes first.
   BIRTH_COHORT: Subjects with birth cohorts between 1942-1962
   CENS: 1 or 0 if the woman has been censored or not (emigration or death before breast cancer event)
3. EVENT: 1 or 0 if the woman has been diagnosed with cancer (given the censoring)
4. AGE1ST: Age at first birth (continuous variable)
5. fruits: Intake of fruits (gram/day) is calculated as apple+banana+juice+orange (A similar variable called fruits and nuts (apple+banana+juice+orange+nuts) has been created for WLH0502. If this variable is used, please note, "nuts" has to be excluded.)
6. vegetables: Intake of vegetables (gram/day) is calculated as broccoli+cabbage+carrot+cauliflower+onion+root+spinach+tomato+bean+pea (A similar variable has already been created for WLH0502. If this variable is used, please note that bean and pea should be added to the variable vegetable within WLH0502.)
7. fruits_veg: Intake of fruits_veg (gram/day) is calculated as the sum of fruits and vegetables

8. fruit_veg_res: Intake of grams fruit and vegetables more or less than expected depending on the estimated energy expenditure. With fruit_veg_res taken as the residuals from the linear regression fruits_veg = a +b*energy.
9. meat: Intake of meats (gram/day) is calculated as the sum of beef, blacksausage, groundmeat, liver, liverpaste, pork, sandwichmeat and sausage. A similar variable (called meat) has been created for WLH0502. This variable may be used but then chicken and game should be removed since in this study we only want to include red meat and not game
10. ALCOHOL: Alcohol intake as categories <5 g, 5-25 g, >25 g per day. 1 drink equals 10 gram. For women who have not filled in the amount of alcohol consumed during the last year but have answered that they do use alcohol the value of 1 gram/day has been imputed.
11. MENSALD: Age at first menstruation (years).
12. OC_USE: Use of oral contraceptives. Categories current, former, never
13. CEDUC: years in school. Categories 0-10, 11-13, >13
14. SMOKING : smoking. Categories: current, former, never
15. AKTDAG: level of physical activity at enrolment. Categories: 1,2,3,4,5
16. VITAMIN_CE: Intake of vitamin C and E (mg/day). This variable is derived as n14+n15..

## *6.2 STATISTICAL ANALYSIS*

### *6.2.1 Plots to understand the relationship between intakes of fruits and vegetables*

Intake of vegetables, fruits, energy, food intake and alcohol have been plotted in order to understand how they are associated (Appendix 2).

### *6.2.2 Censoring*

In the time-to-event analyses the events

- emigration
- death

will be consider censoring events. Subjects with no cancer or censoring event before 31Dec2006 will be considered censored at 31Dec2006.

### *6.2.3 Fruits AND vegetables association with total cancer*

The statistical analyzes will aim at evaluating a qualitative objective, that is ”Is there an association between the intake of fruits and vegetables and total cancer”. This will be done by fitting statistical models and evaluating the support for the different models in the data. The following sets of models/covariates will be considered:

M1. Non-nutritional covariates: Education categories + BMI categories+smoking

M2. (M1) + Alcohol intake (0, <=1, >1 drinks/day) + total energy (kJ/day)

M3. (M2) + fruits_veg_res (gram/day) as continuous *covariate*

M4. (M2) + *quantiles* of fruits_veg_res (gram/day)

M5. (M2) + fruits_veg_res (gram/day) allowing any form *by using splines*

For calculations of fruits_veg residuals fruits_veg will be regressed on total energy intake according to the residual method.

Instead of the standard Cox proportional hazard models Poisson regression models will be fitted since this will facilitate comparisons of the models 1-5 above and allow for a detailed modeling of the time development if needed. The Akaike Information Criteria (AIC) will be calculated for each fitted model. If there is support in the data for an association between cancer and fruits and vegetables the AIC will support one of the models 3-5 in favour of the models 1-2 above.

The AIC will be used to compare the relative importance of fruit and vegetables separately and together: Is there more support in the data for any of these types of food.

For all models the cancer incidence will be evaluated on the attained age as time scale.

Predicted values utilizing model 5 will describe the underlying functional form.

The SAS software procedures GENMOD and TPHREG and the R software (http://cran.r-project.org/) will be used for the statistical analyzes.

### *6.2.4 Fruits association with total cancer*

Repeat 6.2.3 for fruits as exposure in order to investigate the association between fruits only and total cancer. This is done since in this cohort high-consumers of fruit are not the same individuals as high-consumers of vegetables (see Appendix 2).

### *6.2.5 Vegetables association with total cancer*

Repeat 6.2.3 for vegetables as exposure in order to investigate the association between vegetables only and total cancer. This is done since in this cohort high-consumers of fruit are not the same individuals as high-consumers of vegetables (see Appendix 2).

### *6.2.6 Fruits, vegetables and folate and antioxidants*

The underlying biological mechanisms for how fruits and vegetables may protect against cancer are incompletely known. However, the potential cancer protective effects of green, leafy vegetables are often attributed to their folate (one of the B vitamins) content (2). Folate function as a coenzyme in the metabolism of single-carbon compounds for nucleid acid synthesis (part of DNA) and amino acid metabolism. Folate is therefore important for DNA repair (a normal process). Low folate may lead to failed DNA repair which may promote cancer cell growth. Another biological mechanism linking fruits and vegetables to cancer protection is antioxidants(2). Oxidative damage is a part of the cancer process and antixidants ease the oxidative burden. Important antioxidants are vitamin C and E) and fruits and vegetables are rich in these vitamins. Folate, vitamin C and vitamin E are also part of multivitamin supplements.

Therefore, in supplementary models we will try to separate the effect of fruit+vegetables folate and non-folate contribution as well as fruit+vegetables antioxidant and non-antioxidant contribution to the risk of cancer by re-running models M3-M5 above but with folate as well as with vitamin_CE as exposure. In these models use of multivitamin supplement (x_multivitamin: yes or no) will be included as covariate.

# *7 ADDITIONAL ANALYSIS*

## 7.1 Adjustment for physical activity

In the end it will be tested whether the observed associations differ if physical activity (1,2,3,4 or 5) are included as covariate in the final models reported in the manuscript.

## *7.2 Site-specific associations (fruits and vegetables and specific cancers)*

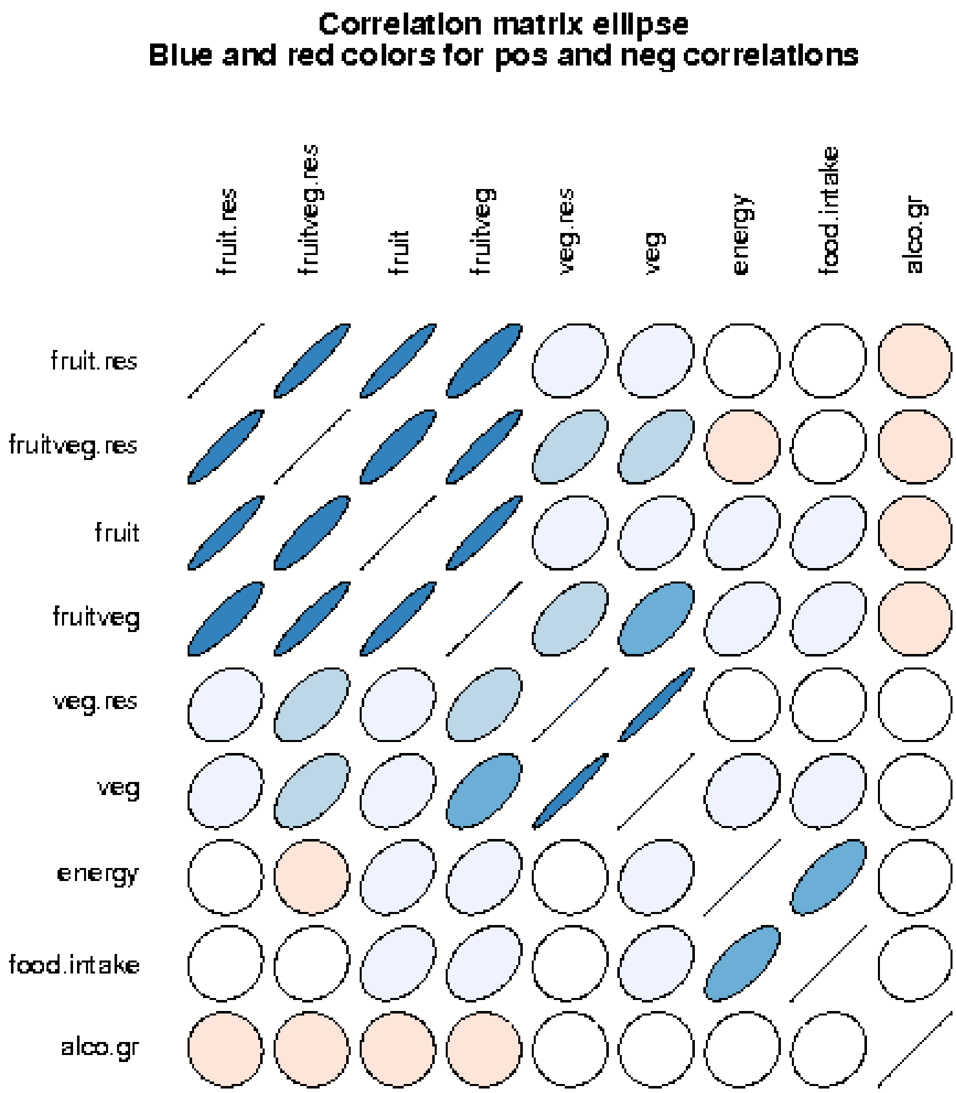
If we do find an association for fruits and vegetables (or fruits or vegetables) we might conduct secondary analysis where we look at different cancer sites. What cancer sites to look at will be discussed and determined at that stage. HO has stated that he wants us to be selective with cancer specific analyses because we know we have limited power- and in many instances, limited originality because the pooling project has published extensively on this, in particular in relation to breast cancer. There might be a stronger case for looking at all gastrointestinal cancer. PL and DT have also suggested that we first go for total cancer first. Later on, they suggest that we can use groups of cancers, e.g. as they have done earlier in Br J Cancer 2008;99:191-5 (smoking-related and non-smoking related cancers). However, what cancer sites to use and how to define them, will be determined later if relevant.

**Appendix 1. ICD-7 codes to include in "total cancer"**

| ICD-7 | Cancer type |
|---|---|
| 170 | Breast |
| 175 | Ovary |
| 193 | Brain and other parts of the nervous system |
| 172 | Corpus uteri |
| 162 | Bronchus and trachea incl lung |
| 153 | Large intestine |
| 154 | Rectum |
| 180 | Kidney |
| 200 | Lymphoscaroma |
| 181 | Bladder and other urinary organs |
| 151 | Stomach |
| 157 | Pancreas |
| 176 | Other unspecified female genital cancer |
| 203 | Multiple myeloma |
| 185 | Bilary passage of the liver |
| 174 | Uterus unspecified |
| 197 | Connective tissue |
| 204 | Leukemia and aleukemia |
| 141 | Tongue |
| 144 | Other parts of the mouth |
| 152 | Small intestine including deodenum |
| 201 | Hodgkins disease |
| 142 | Salivary gland |
| 145 | Oral mesopharnyx |
| 150 | Oesophagus |
| 143 | Floor of mouth |
| 146 | Nasopharynx |
| 160 | Nose |
| 192 | Eye |
| 202 | Other forms of lymphoma |
| 140 | Lip |
| 161 | Larynx |
| 164 | Mediastinum |
| 173 | Other parts of uterus |
| 196 | Bone |
| 206 | Lymphatic system |
| 207 | Hematopoeitic system |

## Appendix 2

Correlations represented as bivariate densities between intakes of vegetables, fruit, alcohol, total food and energy within the WLH cohort. Correlation=0 for independency ==> circle, Correlation=1 for completely dependent==> line.. etc



## References

.

1. Vainio H, Weiderpass E. Fruits and vegetables in cancer prevention. Nutr Cancer 2006;54:111-142.
2. World Cancer Research Fund. Food, nutrition and physical activity and the prevention of cancer. A global perspective, American Institute for Cancer Research, Washington DC, USA, 2007.
3. George SP, Y, Leitzman LF, NF, Dowling ER, J, Schatzkin AH, A, Subar A. Fruit and vegetable intake and risk of cancer: a prospective cohort study. Am J Clin Nutr 2009;89:347-353.
4. Hung H, Joshipura KJ, R et al. Fruit and vegetable intake and risk of major chronic disease. J Natl Cancer Inst 2004;96:1577-1584.
5. Takachi RI, M, Ishihara Jea. Fruit and vegetable intake and risk of total cancer and cardiovascular disease: Japan Public Health Center-based Prospective study. Am J Epidemiology 2008;167:59-70.
6. Benetou V, Orfanos P, Lagiou P, Trichopoulos D, Boffetta P, Trichopoulou A. Vegetables and fruits in relation to cancer risk: evidence from the Greek EPIC Cohort study. Cancer Epidemiol Biomarkers Prev 2008;17:387-392.