Dietary supplement use and risk of colorectal adenomas among persons with Lynch syndrome

MSc Thesis Epidemiology and Public Health

Name:	Esther Swart
Reg. Nr.:	861216819060
Period:	November 2011 – June 2012
Main Supervisor:	Prof. E. kampman, PhD
Daily supervisors:	R. Winkels, PhD
	R. Heine-Bröring, MSc

Preface

When I started working on my thesis, I wrote the following:
"Cancer is one of the major public health problems and a lot of research is going on in this area. Based on the statistics, sooner or later everyone will directly or indirectly suffer from cancer. Cancer really has major consequences for persons who face it. I really would like to be part of the research that tries to reduce cancer rates. I remember a day when I walked through the city and a guy asked me to donate money for some cancer fund. I told him that someday I hoped to do the research on cancer by myself. This thesis might be the start...
My learning objective is to meet my deadlines. Besides, I would like to learn to work in a more structured way. Furthermore, I really would like to apply my statistical knowledge in practice in a real data set using SAS."

Now, after 6 months of working, I can say that this thesis was a good start for me in doing research in cancer epidemiology. I really learned a lot. First of all, by reading about cancer, I gained a lot of new knowledge about cancer. During this thesis I found out that I can work in a structured way. Sometimes I just took my time to structure my files and folders which in the end saved me a lot of time. I really enjoyed working in SAS, and I learned a lot by conducting analyses using SAS. I also learned by experiencing that sometimes things cost more time than you expected on beforehand. I got some delay during the analyses phase of my thesis, which in the end gave me stress when writing the report. I think that facing and handling difficulties that come along with research is a good learning process and therefore, now I finished, I am happy that I experienced some of those difficulties. By entering data and analysing data in a real dataset, I felt doing some very useful things, especially when I read in the hospital notes that persons died because of cancer. There was this moment where I realised that working in cancer epidemiology is what I really want. I really liked being part of the research that tries to find factors in lowering cancer risk.

I would like to thank my supervisors. With the help of Renate and Renate, who planned meetings every 2 weeks, I managed to finish my thesis in time. I think I can still improve in meeting my deadlines, but now I know that with the help of my supervisors I am able to finish in time. My next goal is to try to be not all dependent on deadlines given by my supervisors, but instead really finish things according to my on beforehand made time schedule. I would also thank Renate, Renate and Ellen for all the support, feedback and interesting meetings. They gave me very useful and interesting comments.
I also would like to thank my friends, housemates and thesis buddies Emmy, Nanine, Marlou, Suzanne, Paula, Heleen, Laura, Minke, Anneke, Renee, and Imke for all the nice lunches, tea breaks, good parties, baking cakes and sporting together. Something that I already realised before, but confirmed during this thesis: one should never forget the importance of having breaks. You can struggle a whole morning, but sometimes after a good break, in 5 minutes the problem can be solved!
Very important, I would like to thank my family and the family of my deceased boyfriend. Without their support, I think it would have been a lot harder to start my study again. I am so grateful for that.
Last but not least I would like to thank Lyda, for putting me in room number 109 of the Biotechnion. For me personally, working on this thesis had a really unexpected outcome. I met my current boyfriend in room 109...

This thesis was written in the form of an article. Therefore, additional tables are put in the appendix. The syntax used for all analyses can also be found in the appendix.

Abstract

Purpose: Individuals with Lynch syndrome are at increased risk for colorectal adenomas and carcinomas. Environmental and/or lifestyle factors might affect adenoma risk in Lynch syndrome patients. The objective of this study was to assess whether dietary supplement use modifies the risk of colorectal adenomas in persons with Lynch syndrome.

Methods: This prospective cohort study was conducted among 472 individuals with Lynch syndrome. Cox proportional hazard regression was used to assess hazard ratios (HRs) and 95% confidence intervals (CIs) for the association between dietary supplement use and colorectal adenoma occurrence. To account for dependency of observations within families, robust sandwich estimates of variance was used in analyses. We adjusted for potential confounding by age, sex, smoking habits and number of colonoscopies during person time.

Results: During an overall median follow-up of 28.2 months, 26% (124 out of 472) of the persons in our cohort developed at least one histologically confirmed colorectal adenoma. The intake of any dietary supplement did not statistical significantly increase or decrease the risk of developing colorectal adenomas (HR, 1.21; 95% CI, 0.83 – 1.78). Also for multivitamin intake we did not find a statistical significant association with risk of colorectal adenomas (HR, 1.44; 95% CI, 0.95 – 2.17). Similarly, no statistical significant association was observed with supplemental vitamin C intake and colorectal adenomas (HR, 1.12; 95% CI, 0.72 – 1.75), with supplemental calcium intake and colorectal adenomas (HR 0.95; 95% CI, 0.46 – 1.97), and with other dietary supplements and colorectal adenomas (HR 1.04; 95% CI, 0.59 – 1.84). Intake of fish oil capsules increased the risk of colorectal adenomas with borderline statistical significance (HR 2.39; 95% CI, 1.37 – 4.16).

Conclusion: Dietary supplement intake does not affect risk of developing colorectal adenomas in this cohort of persons with Lynch syndrome.

Table of contents

Introduction

Lynch syndrome, previously named hereditary non-polyposis colorectal cancer (HNPCC), is the most common hereditary syndrome associated with colorectal cancer (CRC) [1]. It is an autosomal dominant disorder accounting for approximately 1-3% of all colorectal cancers [2-6]. Persons with Lynch syndrome are carriers of a germline mutation in one of the mismatch repair (MMR) genes: MLH1, MSH2, MSH6 or PMS2 [1, 7, 8], giving rise to a substantially higher risk of developing CRC. Risk up to age 70 among Lynch syndrome patients ranges from 25 – 70% compared to 2.1 – 2.5% in the general population [9-13].

Both in sporadic CRC and Lynch syndrome, most carcinomas are believed to develop from adenomas, which is supported by the observation that removal of colorectal adenomas is associated with reduced CRC risk [14-17]. However, accelerated carcinogenesis is seen among persons with Lynch syndrome [1]; while it takes 8-10 years for a small adenoma to become a carcinoma in the general population, in Lynch syndrome patients it only takes 2-3 years [1]. Several studies highlight the fact that not all carriers of the MMR gene mutation develop CRC and that age of onset of CRC differs among Lynch syndrome patients [1, 7], suggesting the influence of environmental and/or lifestyle factors.

It is important to determine whether factors affecting sporadic colorectal carcinogenesis, also play a role in adenoma development in individuals with Lynch syndrome [18], because this might lower risk for those individuals at considerable high risk of CRC.

A few studies evaluated associations between diet and/or lifestyle factors in suspected or confirmed carriers of a MMR gene mutation [18-25]. Among the studies performed, results are promising for a possible role of lifestyle factors (e.g. BMI [19, 21], and smoking [20, 23-25]), dietary factors [23] and dietary patterns [22] in modifying risk of developing colorectal adenomas and/or carcinomas.

Evidence for a possible role of dietary supplements in decreasing the risk of colorectal adenomas [26-34] and CRC [35-40] in the general population is inconsistent or inconclusive. Where evidence is suggestive for a protective role for calcium supplementation in developing colorectal adenomas [31-34], the considered precursor of CRC [14-17], for other supplements no clear association was found [26-30]. One study found an even higher risk for some specific supplements [34]. In addition, folic acid, may be involved in cancer progression after the carcinogenic process has started rather than be involved in cancer prevention [41, 42]. Cautiousness is thus suggested for recommendation of supplement intake [29, 33, 35, 37, 41-44]. To our concern, so far, no studies on dietary supplement use among MMR gene mutation carriers are conducted at all.

Individuals living with Lynch syndrome are generally considered to be more health conscious compared to the general population [18], which might result in a higher supplement intake [45]. Considering also the high lifetime risk in developing CRC, it is extremely relevant to prospectively investigate whether dietary supplement use modifies the risk of colorectal adenomas in persons with Lynch syndrome.

Methods

Study population

This prospective cohort study was conducted among individuals with Lynch syndrome of the GEOLynch study. Details of this prospective cohort of patients with Lynch syndrome have been reported previously [21]. In short, eligible MMR gene mutation carriers were identified via the Netherlands Foundation for the Detection of Hereditary Tumours (NFDHT) in Leiden, the Radboud University Nijmegen Medical Centre (RUNMC) in Nijmegen, and the University Medical Centre Groningen (UMCG) in Groningen.

Persons had to be Dutch speaking Caucasian, mentally competent to participate men and women aged between 18 and 80 years, who were screened regularly by colonoscopy to be eligible for our study. Terminally ill patients, and those with familial adenomatous polyposis, inflammatory bowel disease, or a personal history of proctocolectomy or colostomy were excluded.

Between July 2006 and July 2008 a total of 713 known carriers of a germline mutation in at least one of the MMR genes were, with approval of their medical specialist, invited to participate in this study. 686 individuals were eligible for our study. Of these, 73% (499 out of 686) of the persons agreed to participate. Retrieval of medical and/or personal information was not complete for 13 persons. During follow-up, a total of 14 persons did not have one colonoscopy and were therefore excluded from analyses. This resulted in a final cohort of 472 persons. These 472 came from at least 161 families. Approval for this study was obtained from the Medical Ethical committee of the RUNMC. All participants gave written informed consent.

Exposure assessment

At recruitment, dietary supplement intake was assessed using a self-administered questionnaire. Persons could indicate how many times (‘not this month’, ‘once a month’, ‘2-3 days a month’, ‘1 day a week’, ‘2-3 days a week’, ‘4-5 days a week’, and ‘6-7 days a week’) and in what amount they took particularly dietary supplements during the

previous month. Supplements covered in the questionnaire were multivitamins, vitamin C, vitamin B complexes, folic acid, vitamin D (or vitamin A), vitamin E, calcium, iron, and fish oil capsules. In addition, persons could indicate whether they consumed other supplements that were not covered by the questionnaire. A supplement user is defined as a person who took any of the mentioned supplements at least once a month during the previous month. A supplement nonuser is defined as a person who took none of the mentioned supplements during the previous month.

Habitual dietary intake, also of the previous month, was measured using a self-administered, validated 183-item food frequency questionnaire [46, 47]. Information on lifestyle factors was collected using a standardized self-administered questionnaire.

Colorectal adenoma occurrence

Medical information was gathered via the participating centres. From the medical records, information on date, number of colonoscopies and colorectal surgeries, and cancer and adenomatous polyp occurrences before recruitment and during follow-up until December 31, 2010 was extracted.

Detailed information about location, size and histology for all documented polyps that occurred during follow-up was ascertained from pathology reports.

Statistical analysis

In this study, the association between dietary supplement use and the development of colorectal adenomas was assessed among 189 male and 283 female Lynch syndrome patients. Mean age at study entry was 50.4 years. We identified 190 (40%) supplement users. Baseline characteristics of our study population are summarized in table 1 and 2.

Cox proportional hazard regression was used to assess hazard ratios (HRs) and 95% confidence intervals (CIs) for the association between dietary supplement use and colorectal adenoma occurrence. To account for dependency of observations within

families, robust sandwich estimates of variance was used in analyses. The proportional hazard assumption was not violated, as graphically evaluated.

Person-time started at the time of questionnaire completion and ended at the date of diagnosis of the first pathology-confirmed colorectal adenomatous polyp. For persons without an colorectal adenoma diagnosis, person-time was censored at the date of their last known colonoscopy, date of diagnosis of colorectal or extracolonic cancer, date of diagnosis of metastasis, or date of death, whichever occurred first.

We assessed whether the following variables confounded the association between dietary supplement use and colorectal adenomas: age (continuous), sex, educational level (categorical: high versus lower educated), personal history of CRC, personal history of other cancers and personal history of adenomatous polyps (yes/no), number of colonoscopies during person time (categorical: 1, 2, or ≥ 3), BMI (continuous), physical activity level (categorical: high versus lower physically active), smoking habits (categorical: current, former, never), alcohol intake (continuous), NSAID use (more or less than once a month), and fruit intake (continuous). Based on evidence from previous epidemiological research and biological plausibility, we included age, sex, smoking habits, and number of colonoscopies during person time in the basic model. In addition, covariates were included in multivariate models if they were correlated with dietary supplement use and colorectal adenoma development in univariate analyses. Covariates would have remained in the final model when the crude HR changes ≥ 10% after addition of potentially confounding variables using the forward selection method. However, following this criteria, none of the covariates was added to our basic model.

Potential effect modification by sex, history of colorectal neoplasms, smoking habits, and fruit and vegetable intake was examined by stratified analysis and by including interaction terms in multivariate models, where a p for multiplicative interaction less than 0.05 was considered statistically significant. Men and no supplement use, no history of colorectal neoplasms and no supplement use, never smokers and no supplement use, and the lowest quartile of fruit and vegetable intake and no supplement use were set as the reference groups.

Individuals who took any of the mentioned supplements at least once a month during the previous month were considered supplement users. Sensitivity analyses were conducted to assess whether outcomes of our study would have been different if we had defined a supplement user as a person who took any of the mentioned supplements at least once a week during the previous month.

All analyses were performed using SAS, Version 9.2.

Results

During an overall median follow-up of 28.2 months, 26% (124 out of 472) of the persons in our cohort developed at least one histologically confirmed colorectal adenoma (table 1). Compared to the total cohort, cases were slightly older, more likely to be male, lower educated, and were more likely to have had adenomatous polyps in the past. Cases also were more likely to be current or former smokers in comparison to the total cohort. In addition, cases had a higher alcohol intake, a lower fruit intake, and were slightly less likely to use NSAIDs. Any supplement use was slightly higher among the cases, but use of separate individual dietary supplements did not differ between cases and the total cohort (table 1).

190 supplement users (40%) were identified in our cohort. Among the supplement users, 64% (122 out of 190) used multivitamins, 33% (62 out of 190) used vitamin C, 12% (22 out of 190) used calcium, 17% (32 out of 190) used fish oil and 31% (59 out of 190) used supplements not covered by the questionnaire (table 2). The frequency of intake of the other supplements was below 10%. Supplements not covered by the questionnaire mostly were compositions containing glucosamine and garlic capsules. Note that a supplement user was defined as a person who took any of the mentioned supplements at least once a month during the previous month. Some of the persons (20%) took more than one of the various mentioned dietary supplements, and are therefore subsumed under more than one of the individual dietary supplements (table 3).

Compared to nonusers, supplement users were older, more often female, higher educated, more often had a personal history of CRC, other cancers or adenomatous polyps, had a lower alcohol intake, a higher fruit intake, and were more likely to use NSAIDs. While supplement users were less likely to be current and never smokers, the percentage former smokers is higher among the supplement users (table 2).

The intake of any dietary supplement did not statistical significantly increase or decrease the risk of developing colorectal adenomas in this study (HR, 1.21; 95% CI, 0.83 – 1.78). Also for multivitamin intake we did not find a statistical significant association with risk of colorectal adenomas (HR, 1.44; 95% CI, 0.95 – 2.17). Similarly, no statistical

significant association was observed with supplemental vitamin C intake and colorectal adenomas (HR, 1.12; 95% CI, 0.72 – 1.75), with supplemental calcium intake and colorectal adenomas (HR 0.95; 95% CI, 0.46 – 1.97), and with other dietary supplements and colorectal adenomas (HR 1.04; 95% CI, 0.59 – 1.84). Intake of fish oil capsules increased the risk of colorectal adenomas with borderline statistical significance (HR 2.39; 95% CI, 1.37 – 4.16) (table 4).

The association between dietary supplement use and colorectal adenomas was not modified by sex, history of colorectal neoplasms, smoking habits, and fruit and vegetable intake (table 5).

Sensitivity analyses showed that the outcomes of our study would have been the same if a supplement user was defined as a person who took any of the mentioned supplements at least once a week during the previous month (table 6).

Discussion

In this prospective study among 472 persons with Lynch syndrome, after a median follow-up of 28.2 months, any dietary supplement, supplemental multivitamin, supplemental vitamin C, supplemental calcium and other dietary supplements were not statistical significantly associated with risk of colorectal adenomas. For fish oil capsules we did find a borderline statistically significant increased risk of developing colorectal adenomas.

Our results are not completely in line with associations between dietary supplements and colorectal adenomas as found in the general population. We did not find an association between calcium supplement use and colorectal adenomas, where evidence is suggestive but not conclusive for a protective role for calcium supplementation in developing colorectal adenomas in the general population [31-34]. A meta-analysis of eight randomized trials found no convincing association with antioxidant supplements, analysed alone or in combination, and risk of colorectal adenomas in the general population [27], which is in accordance with the result for supplemental vitamin C in our study. Previous studies on multivitamin supplements and fish oil capsule use in the general population focussed on risk of CRC [43, 44, 48, 49]. Results for multivitamin supplements and CRC risk are contradictory [44, 48, 49]. In our study we did not find a statistically significant association with multivitamin supplement use and risk of colorectal adenomas. A statistically significant inverse association with CRC risk was found for fish oil in the general population in the vitamins and lifestyle cohort study [43]. This contradicts with the statistical significant increased risk we found for fish oil on colorectal adenoma development. Note however our small sample size, limiting the statistical power to draw informed conclusions.

A plausible explanation for the heterogeneity in outcomes between studies (i.e. some studies finding an inverse association between supplemental vitamins and colorectal adenomas [31-34], other studies finding no clear association [26-30], and one study

finding an even higher risk for some specific supplements and risk of colorectal adenomas [34]), may be the lack of an uniform assessment of dietary supplement use. Issues concerning dose, form, bioavailability and different combinations and compositions of supplements make it difficult to compare outcomes of different studies using dietary supplement intake as exposure variable [35, 38, 50]. A nutrient found to be protective at a lower dose, may be toxic at a higher dose [35]. In several trials, used doses were substantially above the recommended daily intake, which might eliminate the beneficial effect a supplement might have when taken at a lower dose [38]. Several studies highlight the need to determine how different combinations of supplements affect adenoma and carcinoma development [38, 42]. When several supplements are taken together, or when participants use multivitamins, the possible positive or negative effect of one substance in a supplement, can hypothetically be ruled out by a possible negative or positive effect of another substance in the supplement [39]. Because of the number of supplement users in our cohort, we were only able to show associations for 'any supplement', 'multivitamin', 'vitamin C', 'calcium', 'fish oil', and 'other' intake. It is possible that associations are modified when investigating different supplements simultaneously, which we did for 'any supplement' and 'multivitamin' intake. However, in addition to investigating multiple vitamin supplements simultaneously, we were also able to present associations for various individual supplements (table 4).

Limitations of this study were that we did not take the total number of capsules used at a time into account. In addition, we did not have information about the dosage of supplement intake, which is needed to provide more accurate estimates of dietary supplements and risk of colorectal adenomas and carcinomas [50]. Note however that assessing doses of vitamin supplements might be a time consuming and error-prone task. Currently there are more than 3400 vitamin and mineral preparations available on the market [51], with high inter-brand variability regarding the micronutrient composition of multivitamins and regarding the amount of single vitamins in dietary supplements [50, 52, 53]. Formulations change frequently over time [52, 54, 55], and

label ingredient information not always reflect actual supplement content [54, 55]. For future studies, although extremely difficult to establish, to more accurate assess dietary supplement intake, a supplement composition database might be necessary [50, 52, 56].

Although this is one of the largest prospective cohort studies among confirmed Lynch syndrome cases, the relatively small size of our cohort may have resulted in limited power to detect possible existing associations. Extensive information was available on medical, lifestyle and dietary factors, which made it possible to explore potential confounding of our results in detail. However, residual confounding by unmeasured factors cannot be completely ruled out.

Data on habitual supplement intake in the Netherlands showed that intake in winter was higher than during the rest of the year [57]. Our participants filled out the questionnaire only once at baseline, hence we could not adjust for seasonal variations. 90% (171 out of 190) of our participants filled out the questionnaire during spring or summer (i.e. between 21/3 and 20/9 of each year). This might have caused underreporting of supplement intake. Assuming this kind of misclassification is non-differential, it would probably have biased our results toward the null hypothesis, i.e. no association. However, although only 10% of supplement users (19 out of 190) filled out the questionnaire during autumn or winter (i.e. between 21/9 and 20/3 of each year) intake of supplements in our cohort seemed to be higher in summer than in winter (table 7).

Strengths of this study were a high participation rate of 73% and the inclusion of only confirmed carriers of a mismatch repair gene, making our results generalizable to regularly screened persons with Lynch syndrome in other clinical series. Our study was prospective, which means that information on supplement use was collected before the event of interest. Supplement intake in our cohort was comparable to supplement intake in the general Dutch population [57], indicating that our cohort was not overly health conscious relating the use of dietary supplements.

In conclusion, dietary supplement intake does not affect risk of developing colorectal adenomas in this cohort of persons with Lynch syndrome. As these persons are at high lifetime risk of developing CRC, any observed association between a modifiable factor, including use of dietary supplements, and colorectal adenomas would be of great importance because it allows for recommendations to lower the risk of CRC. Based on results of our study, more, preferably larger studies, collecting accurate information on the dose of dietary supplements, are highly needed before clear advices regarding the use of dietary supplements for persons with Lynch syndrome can be given. In the meantime, cautiousness is suggested for supplement intake.

References

1. Cunningham, D., *Colorectal cancer.* Lancet, 2010. **375**: p. 1030-1047.
2. Salovaara, R., *Population-Based Molecular Detection of Hereditary Nonpolyposis Colorectal Cancer.* J Clin Oncol, 2000. **18**(11): p. 2193-2200.
3. Hampel, H., *Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer).* N Enlg J Med, 2005. **352**(18): p. 1851-1860.
4. Piñol, V., *Accuracy of Revised Bethesda Guidelines, Microsatellite Instability, and Immunohistochemistry for the Identification of Patients With Hereditary Nonpolyposis Colorectal Cancer.* JAMA, 2005. **293**(16): p. 1986-1994.
5. *National Cancer Institute: What you need to know about cancer of the colon and rectum*. [cited 2011 12-12]; Available from: http://www.cancer.gov/cancertopics/wyntk/colon-and-rectal/page4
6. Barnetson, R., *Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer.* N Engl J Med, 2006. **354**(26): p. 2751-2763.
7. Vasen, H.F.A., *Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer).* J Med Genet, 2007. **44**: p. 353-362.
8. Foulkes, W., *Inherited susceptibility to common cancers.* N Engl J Med, 2008. **359**: p. 2143-2153.
9. Barrow, E., *Colorectal cancer in HNPCC: cumulative lifetime incidence, survival and tumour distribution. A report of 121 families with proven mutations.* Clinical Genetics, 2008. **74**(3): p. 233-242.
10. Quehenberger, F., *Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment.* J Med Genet, 2005. **42**: p. 491-496.
11. Aarnio, M., et al., *Cancer risk in mutation carriers of DNA-mismatch-repair genes.* International Journal of Cancer, 1999. **81**(2): p. 214-218.

12. Jenkins, M.A., et al., *Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study.* Clinical Gastroenterology and Hepatology, 2006. **4**(4): p. 489-498.
13. Kankerbestrijding, S.K.v.K., *Signaleringsrapport 'De kans op kanker. Bewerking van cijfers NKR en CBS 1999-2003'*2007, Amsterdam: Interne rapportage KWF Kankerbestrijding.
14. Atkin, W.S., B.C. Morson, and J. Cuzick, *Long-term risk of colorectal cancer after excision of rectosigmoid adenomas.* New England Journal of Medicine, 1992. **326**(10): p. 658-662.
15. Järvinen, H.J., et al., *Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer.* Gastroenterology, 2000. **118**(5): p. 829-834.
16. Vasen, H.F.A., et al., *One to 2-year surveillance intervals reduce risk of colorectal cancer in families with Lynch syndrome.* Gastroenterology, 2010. **138**(7): p. 2300-2306.
17. Winawer, S.J., et al., *Prevention of colorectal cancer by colonoscopic polypectomy.* New England Journal of Medicine, 1993. **329**(27): p. 1977-1981.
18. Voskuil, D., et al., *Meat consumption and meat preparation in relation to colorectal adenomas among sporadic and HNPCC family patients in The Netherlands.* European Journal of Cancer, 2002. **38**(17): p. 2300-2308.
19. Campbell, P.T., et al., *Excess body weight and colorectal cancer risk in Canada: associations in subgroups of clinically defined familial risk of cancer.* Cancer Epidemiology Biomarkers & Prevention, 2007. **16**(9): p. 1735-1744.
20. Pande, M., et al., *Smoking and colorectal cancer in Lynch syndrome: results from the Colon Cancer Family Registry and the University of Texas MD Anderson Cancer Center.* Clinical Cancer Research, 2010. **16**(4): p. 1331.
21. Botma, A., et al., *Body mass index increases risk of colorectal adenomas in men with Lynch syndrome: the GEOLynch cohort study.* Journal of Clinical Oncology, 2010. **28**(28): p. 4346-4353.

22. Botma, A., et al., *Dietary patterns and colorectal adenomas in Lynch syndrome: The GEOLynch cohort study.* Submitted.

23. Diergaarde, B., et al., *Environmental factors and colorectal tumor risk in individuals with hereditary nonpolyposis colorectal cancer.* Clinical Gastroenterology and Hepatology, 2007. **5**(6): p. 736-742. e1.

24. Watson, P., et al., *Tobacco use and increased colorectal cancer risk in patients with hereditary nonpolyposis colorectal cancer (Lynch syndrome).* Archives of internal medicine, 2004. **164**(22): p. 2429.

25. Winkels, R.M., et al., *Smoking Increases the Risk for Colorectal Adenomas in Persons with Lynch Syndrome.* Gastroenterology, 2011.

26. Greenberg, E.R., et al., *A clinical trial of antioxidant vitamins to prevent colorectal adenoma.* New England Journal of Medicine, 1994. **331**(3): p. 141-147.

27. Bjelakovic, G., et al., *Meta-analysis: antioxidant supplements for primary and secondary prevention of colorectal adenoma.* Alimentary pharmacology & therapeutics, 2006. **24**(2): p. 281-291.

28. Wu, K., et al., *A randomized trial on folic acid supplementation and risk of recurrent colorectal adenoma.* The American journal of clinical nutrition, 2009. **90**(6): p. 1623-1631.

29. Cole, B.F., et al., *Folic acid for the prevention of colorectal adenomas.* JAMA: the journal of the American Medical Association, 2007. **297**(21): p. 2351.

30. Figueiredo, J.C., et al., *Folic acid and prevention of colorectal adenomas: A combined analysis of randomized clinical trials.* International Journal of Cancer, 2010.

31. Peters, U., et al., *Calcium intake and colorectal adenoma in a US colorectal cancer early detection program.* The American journal of clinical nutrition, 2004. **80**(5): p. 1358-1365.

32. Baron, J., et al., *Calcium supplements for the prevention of colorectal adenomas.* New England Journal of Medicine, 1999. **340**(2): p. 101-107.

33. Weingarten, M.A., A. Zalmanovici, and J. Yaphe, *Dietary calcium supplementation for preventing colorectal cancer and adenomatous polyps.* Cochrane Database Syst Rev, 2008. **1**.

34. Bonithon-Kopp, C., et al., *Calcium and fibre supplementation in prevention of colorectal adenoma recurrence: a randomised intervention trial.* The Lancet, 2000. **356**(9238): p. 1300-1306.

35. WCRF/AICR, *Food, nutrition, physical activity, and the prevention of cancer: a global perspective*. 2nd ed2007, Washington DC, AICR.

36. WCRF/AICR, *Continuous update project interim report summary: food, nutrition and physical activity and the prevention of colorectal cancer*2011.

37. Wu, K., et al., *A prospective study on supplemental vitamin e intake and risk of colon cancer in women and men.* Cancer Epidemiology Biomarkers & Prevention, 2002. **11**(11): p. 1298-1304.

38. Greenwald, P., et al., *Clinical trials of vitamin and mineral supplements for cancer prevention.* The American journal of clinical nutrition, 2007. **85**(1): p. 314S-317S.

39. Bjelakovic, G., et al., *Antioxidant supplements for preventing gastrointestinal cancers.* Cochrane Database Syst Rev, 2008. **3**.

40. Wactawski-Wende, J., et al., *Calcium plus vitamin D supplementation and the risk of colorectal cancer.* New England Journal of Medicine, 2006. **354**(7): p. 684-696.

41. Ulrich, C.M. and J.D. Potter, *Folate supplementation: Too much of a good thing?* Cancer Epidemiology Biomarkers & Prevention, 2006. **15**(2): p. 189-193.

42. Ulrich, C.M. and J.D. Potter, *Folate and cancer—timing is everything.* JAMA: the journal of the American Medical Association, 2007. **297**(21): p. 2408.

43. Satia, J.A., et al., *Associations of herbal and specialty supplements with lung and colorectal cancer risk in the VITamins and Lifestyle study.* Cancer Epidemiology Biomarkers & Prevention, 2009. **18**(5): p. 1419-1428.

44. White, E., J.S. Shannon, and R.E. Patterson, *Relationship between vitamin and calcium supplement use and colon cancer.* Cancer Epidemiology Biomarkers & Prevention, 1997. **6**(10): p. 769-774.

45. Touvier, M., et al., *Dietary and cancer–related behaviors of vitamin/mineral dietary supplement users in a large cohort of French women.* European journal of nutrition, 2006. **45**(4): p. 205-214.

46. Feunekes, G.I., et al., *Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol.* The American journal of clinical nutrition, 1993. **58**(4): p. 489-496.

47. Verkleij-Hagoort, A.C., et al., *Validation of the assessment of folate and vitamin B12 intake in women of reproductive age: the method of triads.* European journal of clinical nutrition, 2006. **61**(5): p. 610-615.

48. Park, Y., *Intakes of vitamins A, C, and E and use of multiple vitamin supplements and risk of colon cancer: a pooled analysis of prespective cohort studies.* Cancer Causes Control, 2010. **21**: p. 1745-1757.

49. White, E., et al., *VITamins And Lifestyle cohort study: study design and characteristics of supplement users.* American journal of epidemiology, 2004. **159**(1): p. 83.

50. Ocke, M.C., E. Buurma-Rethans, and H. Fransen, *Dietary supplement use in the Netherlands: current data and recommendations for future assessment.* RIVM Report 2005. **350100001**.

51. Park, Y.K., I. Kim, and E.A. Yetley, *Characteristics of vitamin and mineral supplement products in the United States.* The American journal of clinical nutrition, 1991. **54**(4): p. 750-759.

52. Patterson, R.E., et al., *Validity of methods used to assess vitamin and mineral supplement use.* American journal of epidemiology, 1998. **148**(7): p. 643-649.

53. Van Kreijl, C. and A. Knaap, *Ons eten gemeten: Gezonde voeding en veilig voedsel in Nederland*2007: Bohn Stafleu van Loghum.

54. Hendrich, S. and K. Fisher, *What do we need to know about active ingredients in dietary supplements? Summary of workshop discussion.* The Journal of nutrition, 2001. **131**(4): p. 1387S-1388S.

55. Yetley, E.A., *Multivitamin and multimineral dietary supplements: definitions, characterization, bioavailability, and drug interactions.* The American journal of clinical nutrition, 2007. **85**(1): p. 269S-276S.

56. Patterson, R.E., et al., *Evaluation of a simplified vitamin supplement inventory developed for the Women's Health Initiative.* Public health nutrition, 1999. **2**(03): p. 273-276.

57. Rossum, C.v., *Dutch National Food Consumption Survey 2007-2010 : Diet of children and adults aged 7 to 69 years* 2011. p. 3-135.

Tables

Table 1. Baseline characteristics of the 124 adenomatous polyp cases in comparison to the total cohort of 472 persons with Lynch syndrome

Characteristic	Adenomatous polyp cases (n=124)	Total cohort (n=472)
Person Months [median (IQR[a)])]	16.9 (7.9 – 32.7)	28.2 (17.1 – 39.3)
Demographic characteristics		
Age at study entry, years [median (IQR[a)])]	52.7 (45.6 – 59.9)	50.4 (40.8 – 58.7)
Sex, male [n (%)]	54 (44)	189 (40)
Education level, higher [n (%)][b)]	29 (24)	160 (34)
Medical characteristics [n(%)]		
MMR gene mutation		
MLH1	49 (40)	180 (38)
MSH2	56 (45)	192 (41)
MSH6	18 (15)	95 (20)
PMS2	1 (1)	3 (1)
History of colorectal cancer	31 (25)	123 (26)
History of adenomatous polyps	58 (47)	150 (32)
History of other cancers	26 (21)	84 (18)
No. of colonoscopies during person time [n (%)]		
1	64 (52)	113 (24)
2	25 (20)	146 (31)
≥3	35 (28)	213 (45)
Colon surgery [n (%)]		
None	89 (72)	338 (72)
Partial colon resection	30 (24)	80 (17)
Subtotal colectomy	4 (3)	41 (9)
Supplement intake [n (%)][c)]		
Any use, yes [d)]	55 (44)	190 (40)
Multivitamins	37 (30)	122 (26)

Vitamin C intake	17 (14)	62 (13)
Vitamin B complexes	1 (1)	8 (2)
Folic acid	2 (2)	9 (2)
Vitamin D (or vitamin A)	1 (1)	6 (1)
Vitamin E	1 (1)	9 (2)
Calcium	6 (5)	22 (5)
Iron	0 (0)	4 (1)
Fish oil capsules	13 (10)	32 (7)
Other[e)]	20 (16)	59 (13)
Lifestyle and dietary factors		
BMI, kg/m^2 [median (IQR[a)])]	25.1 (23.4 – 27.1)	24.5 (22.5 – 27.0)
Height, cm [median (IQR[a)])]	174 (167 - 180)	174 (167 – 180)
Physical activity, high [n (%)][f)]	42 (35)	156 (34)
Smoking habits [n (%)]		
Current	36 (29)	86 (18)
Former	62 (50)	204 (43)
Never	26 (21)	181 (38)
Alcohol intake, g/d [median (IQR[a)])]	9.4 (3.2 – 21.7)	7.2 (1.5 – 16.8)
NSAID use ≥ 1x/month [n (%)]	23 (19)	98 (21)
Total dietary energy intake, kcal/d [median (IQR[a)])]	2003 (1608 – 2488)	2059 (1706– 2550)
Total fibre intake, g/d [median (IQR[a)])]	23.6 (16.8 – 30.9)	25.2 (19.8 – 31.0)
Vegetable intake, g/d [median (IQR[a)])]	124 (62 – 173)	123 (76 – 175)
Fruit intake, g/d [median (IQR[a)])]	136 (44 – 235)	156 (75 – 234)
Red meat intake, g/d [median (IQR[a)])]	49 (30 – 67)	46 (30 – 64)
Processed meat intake g/d [median (IQR[a)])]	11 (4 – 26)	12 (4 – 27)

Abbreviation: n, number; IQR, inter quartile range; MMR, mismatch repair; wk, week; BMI, body mass index; kg, kilo gram; m, meter; cm, centimeter; kcal, kilocalories; d, day; g, gram; NSAID, non-steroidal anti-inflammatory drugs

Notes:
a) IQR inter quartile range is the 25th – 75th percentile
b) Higher education is a college or university education
c) Supplement user is defined as a person who took any of the mentioned supplements at least once a month during the previous month
d) Is yes if a person takes any of the mentioned individual supplements
e) Supplements not covered by the questionnaire, mostly compositions containing glucosamine and garlic capsules
f) High physical activity is the highest tertile of the physical activity score

Table 2. Baseline characteristics of the 472 persons with Lynch syndrome, stratified by supplement use

Characteristic	Supplement user[c)] (n=190)	Supplement nonuser (n=282)
Person Months [median (IQR[a)])]	27.1 (16.6 – 39.7)	28.8 (17.7 – 39.2)
Demographic characteristics		
Age at study entry, years [median (IQR[a)])]	51.7 (42.1 – 59.2)	48.9 (39.6 – 58.8)
Sex, male [n (%)]	60 (32)	129 (46)
Education level, higher [n (%)][b)]	70 (37)	90 (32)
Medical characteristics [n(%)]		
MMR gene mutation		
MLH1	78 (41)	102 (36)
MSH2	66 (35)	126 (45)
MSH6	43 (23)	52 (18)
PMS2	2 (1)	1 (0)
History of colorectal cancer	52 (27)	71 (25)
History of adenomatous polyps	65 (34)	85 (30)
History of other cancers	43 (23)	41 (15)
No. of colonoscopies during person time [n (%)]		
1	46 (24)	67 (24)
2	57 (30)	89 (32)
≥3	87 (46)	126 (45)
Colon surgery [n (%)]		
None	128 (67)	210 (74)
Partial colon resection	40 (21)	40 (15)
Subtotal colectomy	20 (11)	21 (8)
Supplement intake [n (%)][c)]		
Multivitamins	122 (64)	-
Vitamin C intake	62 (33)	-
Vitamin B complexes	8 (4)	-
Folic acid	9 (5)	-

Vitamin D (or vitamin A)	6 (3)	-
Vitamin E	9 (5)	-
Calcium	22 (12)	-
Iron	4 (2)	-
Fish oil capsules	32 (17)	-
Other[d)]	59 (31)	-
Lifestyle and dietary factors		
BMI, kg/m^2 [median (IQR[a)])]	24.1 (21.9 – 26.4)	24.7 (22.9 – 27.5)
Height, cm [median (IQR[a)])]	174 (167 – 180)	174 (167 – 180)
Physical activity, high [n (%)][e)]	65 (35)	91 (33)
Smoking habits [n (%)]		
Current	32 (17)	54 (19)
Former	93 (49)	111 (39)
Never	65 (34)	116 (41)
Alcohol intake, g/d [median (IQR[a)])]	6.4 (1.5 – 16.1)	7.4 (1.5 – 17.1)
NSAID use ≥ 1x/month [n (%)]	42 (23)	56 (20)
Total dietary energy intake, kcal/d [median (IQR[a)])]	1986 (1711 – 2526)	2089 (1699 – 2568)
Total fibre intake, g/d [median (IQR[a)])]	25 (21 – 31)	25 (19 – 31)
Vegetable intake, g/d [median (IQR[a)])]	122 (81 – 176)	125 (73 – 175)
Fruit intake, g/d [median (IQR[a)])]	166 (79 – 235)	130 (53 – 233)
Red meat intake, g/d [median (IQR[a)])]	46 (27 – 63)	46 (34 – 64)
Processed meat intake g/d [median (IQR[a)])]	12 (3 – 23)	12 (6 – 31)

Abbreviation: n, number; IQR, inter quartile range; MMR, mismatch repair; wk, week; BMI, body mass index; kg, kilo gram; m, meter; cm, centimeter; kcal, kilocalories; d, day; g, gram; NSAID, non-steroidal anti-inflammatory drugs

Notes:
a) IQR inter quartile range is the 25th – 75th percentile
b) Higher education is a college or university education
c) Supplement user is defined as a person who took any of the mentioned supplements at least once a month during the previous month
d) Supplements not covered by the questionnaire, mostly compositions containing glucosamine and garlic capsules
e) High physical activity is the highest tertile of the physical activity score

Table 3. Overview of participants taking different dietary supplements simultaneously

Amount of different dietary supplements taken simultaneously	**Persons frequency (%) (total n=472)**
0	282 (60)
1	97 (21)
2	59 (13)
3	22 (5)
4	8 (2)
5	4 (1)

Table 4. Hazard ratios for supplement use (≥ 1 x/month) and development of colorectal adenomatous polyps in a cohort of 472 persons with Lynch syndrome

Dietary supplement	**No use (n=282)**	**Use (n=190)**
	HR	**HR (95% CI)**
Any dietary supplement		
Person time[a)] [median (IQR[b)])]	16.92 (7.13 – 30.65)	16.85 (8.15 – 34.17)
No. of cases	69	55
HR[c)]	1.0	1.21 (0.83 – 1.78)
Multivitamins		
Person time[a)] [median (IQR[b)])]	16.20 (6.57 – 30.65)	19.29 (11.76 – 34.46)
No. of cases	87	37
HR[c)]	1.0	1.44 (0.95 – 2.17)
Vitamin C		
Person time[a)] [median (IQR[b)])]	16.92 (7.82 – 33.05)	16.53 (11.76 – 28.68)
No. of cases	107	17
HR[c)]	1.0	1.12 (0.72 – 1.75)
Calcium		
Person time[a)] [median (IQR[b)])]	16.69 (7.13 – 32.26)	28.80 (15.38 – 34.04)
No. of cases	118	6
HR[c)]	1.0	0.95 (0.46 – 1.97)
Fish oil		
Person time[a)] [median (IQR[b)])]	16.53 (7.13 – 30.98)	29.96 (9.07 – 34.46)
No. of cases	111	13
HR[c)]	1.0	2.39 (1.37 – 4.16)
Other		
Person time[a)] [median (IQR[b)])]	16.89 (8.02 – 30.60)	16.25 (6.32 – 36.63)
No. of cases	104	20
HR[c)]	1.0	1.04 (0.59 – 1.84)

Notes
a) Person months
b) IQR inter quartile range is the 25^{th} – 75^{th} percentile
c) Adjusted for age, sex, smoking habits, and number of colonoscopies during person time

Table 5. Testing for potential effect modification by sex, history of colorectal neoplasms, smoking status, and fruit and vegetable intake

Table 5.1. Stratification by sex

	Sex	
Dietary supplement use	**Men**	**Women**
No use		
No. of cases	35	34
HR (95% CI)	1.0	0.85 (0.52 – 1.41)
Use		
No. of cases	19	36
HR (95% CI)	1.16 (0.70 – 1.93)	1.06 (0.66 – 1.71)
P for interaction		0.83

Table 5.2. Stratification by history of colorectal neoplasms

	History of colorectal neoplasm	
Dietary supplement use	**No**	**Yes**
No use		
No. of cases	32	37
HR (95% CI)	1.0	1.00 (0.60 – 1.68)
Use		
No. of cases	19	36
HR (95% CI)	1.07 (0.55 – 2.09)	1.30 (0.80 – 2.11)
P for interaction		0.63

Table 5.3. Stratification by smoking status

	Smoking status		
Dietary supplement use	**Never**	**Former**	**Current**
No use			
No. of cases	17	29	23
HR (95% CI)	1.0	1.70 (0.86 – 3.33)	3.62 (1.86 – 7.02)
Use			
No. of cases	9	33	13
HR (95% CI)	0.86 (0.37 – 1.99)	2.88 (1.52 – 5.49)	3.12 (1.59 – 6.10)
P for interaction			0.88

Table 5.4. Stratification by fruit and vegetable intake

	Total fruitveg intake (g/day)			
Dietary supplement use	**Q1 < 179**	**Q2 179 - 283**	**Q3 283 - 387**	**Q4 > 387**
No use				
No. of cases	26	10	15	18
HR (95% CI)	1.0	0.69 (0.32 – 1.48)	1.03 (0.55 – 1.91)	1.02 (0.54 – 1.95)
Use				
No. of cases	16	11	14	14
HR (95% CI)	1.08 (0.56 – 2.07)	0.91 (0.40 – 2.09)	1.16 (0.57 – 2.35)	1.56 (0.75 – 3.27)
P for interaction				0.51

Table 6. Hazard ratios for supplement use (≥ 1 x/wk) and development of colorectal adenomatous polyps in a cohort of 472 persons with Lynch syndrome

Dietary supplement	**No use (n=299)**	**Use (n=173)**
	HR	**HR (95% CI)**
Any dietary supplement		
No. of cases	75	49
HR[a)]	1.0	1.13 (0.77 – 1.66)
Multivitamins		
No. of cases	92	32
HR[a)]	1.0	1.39 (0.90 – 2.14)
Vitamin C		
No. of cases	111	13
HR[a)]	1.0	1.00 (0.61 – 1.66)
Calcium		
No. of cases	118	6
HR[a)]	1.0	1.04 (0.51 – 2.14)
Fish oil		
No. of cases	112	12
HR[a)]	1.0	2.32 (1.29 – 4.17)
Other		
No. of cases	106	18
HR[a)]	1.0	0.98 0.53 – 1.80)

a) Adjusted for age, sex, smoking habits, and number of colonoscopies during person time

Table 7. Use per season of interview

Dietary supplement	Winter (%) (n=57)	Summer (%) (n=415)
Any use, yes[a)]	19 (33)	171 (41)
Multivitamins	15 (26)	107 (26)
Vitamin C intake	9 (16)	53 (13)
Vitamin B complexes	1 (2)	7 (2)
Folic acid	0 (0)	9 (2)
Vitamin D (or vitamin A)	0 (0)	6 (1)
Vitamin E	1 (2)	8 (2)
Calcium	0 (0)	22 (5)
Iron	0 (0)	4 (1)
Fish oil capsules	1 (2)	31 (7)
Other	5 (9)	54 (13)

a) Is yes if a person takes any of the mentioned individual supplements

Appendices

Table A1. Additional baseline characteristics of the 70 female adenomatous polyp cases in comparison to the total female cohort of 283 women with Lynch syndrome

Characteristic	Adenomatous polyp cases (n=70)	Total cohort (n=283)
Menopausal status, yes [n (%)][a)]	47 (67)	157 (55)
Postmenopausal hormone use, yes [n (%)][b)]	15 (32)	37 (24)

a) Of the total female cohort

b) Among menopausal women

Table A2. Additional baseline characteristics of the 283 women with Lynch syndrome, stratified by supplement use

Characteristic	Supplement user (n=130)	Supplement nonuser (n=153)
Menopausal status, yes [n (%)]	81 (62)	76 (50)
Postmenopausal hormone use, yes [n (%)][a)]	24 (30)	13 (17)

a) Among menopausal women

Table A3. Hazard ratios for supplement use (≥ 1 x/month) and development of colorectal adenomatous polyps in a cohort of 283 women with Lynch syndrome

Dietary supplement	No use (n=153) HR	Use (n=130) HR (95% CI)
Any dietary supplement		
No. of cases	34	36
HR[a)]	1.0	1.21 (0.72 – 2.01)
HR[b)]	1.0	1.24 (0.74 – 2.10)

a) Adjusted for for age, sex, smoking habits, and number of colonoscopies during person time

b) Additionally adjusted for menopausal status and postmenopausal hormone use

Table A4. Average length of our cohort of 472 persons with Lynch syndrome

Gender	**Average length [median (IQR[a)])]**
Men	182 (177 – 187)
Women	169 (163 – 174)

a) IQR inter quartile range is the 25th – 75th percentile

Syntax used for analyses

```
LIBNAME ES 'M:Thesis2';

* Change (add) variable to merge ffqtotal, Geolynch_source, extravar_es and
Extravariabelen_renate, see merge step;
DATA Geolynch_source2;
SET ES.Geolynch_source;
deelnr=respnr;
RUN;

DATA Extravariabelen_renate2;
SET ES.Extravariabelen_renate;
deelnr=respnr;
RUN;

DATA extravar_es2;
SET ES.extravar_es;
deelnr=respnr;
RUN;

* PROC SORT because want to use the by statement to merge;
PROC SORT DATA=Geolynch_source2;
BY deelnr;
RUN;

PROC SORT DATA=ES.ffqtotal;
BY deelnr;
RUN;

PROC SORT DATA=Extravariabelen_renate2;
BY deelnr;
RUN;

PROC SORT DATA=extravar_es2;
BY deelnr;
RUN;

* Combine ffqtotal, Geolynch_source2, extravar_es and
Extravariabelen_renate2 for all variables;
DATA total;
MERGE ES.ffqtotal Geolynch_source2 extravar_es2 extravariabelen_renate2;
BY deelnr;
RUN;

* To check of variabelen overeenkomen met art A.B. en wat labeling inhoud;
* Variable is at study entry;
```

```
PROC SORT DATA=total;
BY CRTHist;
RUN;

* H2 A.B.;
PROC FREQ DATA=total;
TABLES operatie dumoperatiepart dumoperatiesubtot;
BY CRThist;
RUN;

* H3 A.B.;
PROC FREQ DATA=total;
TABLES operatie dumoperatiepart dumoperatiesubtot;
RUN;

* Create variable BMI;
DATA total1;
SET total;
BMI = weight/(length/100)**2;
RUN;

* Create meatvariables, see next step;
DATA total2;
SET total1;
totmincedmeat = (0.1 * meatunknown) + mincedmeat;
totbeeflowfat =   (0.1 * meatunknown) + beeflowfat;
totbeeffat = (0.1 * meatunknown) + beeffat;
totporklowfat =   (0.1 * meatunknown) + porklowfat;
totporkmedium =   (0.1 * meatunknown) + porkmedium;
totporkfat = (0.1 * meatunknown) + porkfat;
totmeatoth = (0.1 * meatunknown) + meatoth;

totproccessedmeatlowfat = (0.166666667 * proccessedmeatunkn) +
proccessedmeatlowfat;
totproccessedmeatliver = (0.166666667 * proccessedmeatunkn) +
proccessedmeatliver;
totproccessedmeatmedium = (0.166666667 * proccessedmeatunkn) +
proccessedmeatmedium;
totproccessedmeatfat = (0.166666667 * proccessedmeatunkn) +
proccessedmeatfat;
RUN;

* Create variable alltotveggies, alltotfruit, redmeat and processed meat;
DATA total3;
SET total2;
alltotveggies = cauliflower + broccoli + cabbageraw + spinach + lettuce +
carrot + carrotraw + tomato +
allium + veggieoth + rawveggieoth;
```

```
alltotfruit = orange + apple + pear + banana + strawberry + grape +
fruitoth;
redmeat = totmincedmeat + totbeeflowfat + totbeeffat + totporklowfat +
totporkmedium
+ totporkfat + totmeatoth;
alltotprocessedmeat = totproccessedmeatlowfat + totproccessedmeatliver +
totproccessedmeatmedium
+ totproccessedmeatfat + sausageliver + sausageoth;
RUN;

PROC PRINT DATA=total3;
VAR deelnr START;
RUN;

* Een getal maken van START variabele, om seizoen aan te maken, zie '2012-
04-19 meeting7.doc'.
Wat is 21feb2006, 20aug2006, etc? want dat zijn de season cut off data
points. Deze precieze (cut off) data waardes
komen niet onder START variabele voor, dus eerst in ander bestand (via
excel en SPSS naar SAS)
deze data aangemaakt om cutoffpoints te bepalen;
DATA seasoncutoff1;
SET ES.seasoncutoff;
datenumber=cutoff;
RUN;

* Getal maken voor alle data waardes onder START variabele;
DATA total3_1;
SET total3;
STARTnumber=START;
RUN;

* Seasonvariable aanmaken. Voor season variable, season period 1 maand naar
voren geschoven,
want ffq gaat over periode maand voor START variable.
Onderstaande waardes komen uit dataset seasoncutoff1;
DATA total3_2;
SET total3_1;
IF STARTnumber >= 16853 AND STARTnumber <= 17033 THEN season='SpSu';
IF STARTnumber >= 17034 AND STARTnumber <= 17217 THEN season='AuWi';
IF STARTnumber >= 17218 AND STARTnumber <= 17398 THEN season='SpSu';
IF STARTnumber >= 17399 AND STARTnumber <= 17582 THEN season='AuWi';
IF STARTnumber >= 17583 AND STARTnumber <= 17764 THEN season='SpSu';
IF STARTnumber >= 17765 AND STARTnumber <= 17948 THEN season='AuWi';
IF STARTnumber >= 17949 AND STARTnumber <= 18129 THEN season='SpSu';
IF STARTnumber >= 18130 AND STARTnumber <= 18313 THEN season='AuWi';
RUN;

* Check seasonvariable, klopt;
```

```
PROC PRINT DATA=total3_2;
VAR START STARTnumber season;
RUN;

* Datacleaning;
DATA total4;
SET Total3_2;
IF respnr= 'A1098' THEN v86= 1;
IF respnr= 'A1302' THEN v86= '.';
IF respnr= 'A1313' THEN v86= 1;
IF respnr= 'A1389' THEN v86= 1;
IF respnr= 'A1416' THEN v86= 1;
IF respnr= 'A1438' THEN v86= 1;
IF respnr= 'A1445' THEN v86= 1;
IF respnr= 'A1449' THEN v86= 1;
IF respnr= 'A1451' THEN v86= 1;
IF respnr= 'A1456' THEN v86= 1;
IF respnr= 'A1481' THEN v86= 1;
IF respnr= 'A1483' THEN v86= 1;
IF respnr= 'A1485' THEN v86= 1;
IF respnr= 'A1486' THEN v86= 1;
IF respnr= 'A1490' THEN v86= 1;
IF respnr= 'A2052' THEN v86= 1;
IF respnr= 'A2055' THEN v86= 1;
IF respnr= 'A2057' THEN v86= 1;
IF respnr= 'A2064' THEN v86= 1;
IF respnr= 'A2066' THEN v86= 1;
IF respnr= 'A2083' THEN v86= 1;
IF respnr= 'A2113' THEN v86= 1;
RUN;

* Change (add) variable to merge total4 and Fu2comleet2_final_restructured,
see next step;
DATA Fu2compleet2_final_restructured2;
SET ES.Fu2compleet2_final_restructured1;
deelnr=respondentnummer;
RUN;

* PROC SORT because want to use the by statement to merge;
PROC SORT DATA=Fu2compleet2_final_restructured2;
BY deelnr;
RUN;

PROC SORT DATA=total4;
BY deelnr;
RUN;

* Combine total4 and Fu2compleet2_final_restructured2 for all variables;
DATA total5;
```

```
MERGE total4 Fu2compleet2_final_restructured2;
BY deelnr;
RUN;

* Om in A1437, A2118 en A2120 data aan te passen, zie syntax hieronder;
DATA total5_getaldatum;
SET total5;
getal_dtscltst_1=dtscltst_1;
RUN;

PROC PRINT DATA=total5_getaldatum;
VAR deelnr STOETnr dtscltst_1 getal_dtscltst_1;
RUN;

/* Invoeren missende scopieen van de StOET */

DATA total55;
SET total5;

IF deelnr='A1002' THEN DO;
stoetnr=1530012; dtscltst_1=18323; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;
FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1029' THEN DO;
stoetnr=1100033; dtscltst_1=18239; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

dtscltst_2=18609; kanker_2='Nee'; coecumwur_2=2; srtscr_2=1;
lnc_clnscp_2=99;
clean_2=1; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1065' THEN DO;
stoetnr=89038; dtscltst_1=18212; kanker_1='Nee'; coecumwur_1=1; srtscr_1=1;
clean_1=1; scopie_1=7; poliepaantal_1=1; npoliep_1=2; grpolsc1_1=99;
grpolpa1_1=99;
polpect_1=3; PArap_1=1; srtpoliep1_1=1; locp1_1=7;  grpolsc2_1=99;
grpolpa2_1=99;
srtpoliep2_1=1; locp2_1=7; advice_1=1; adviceyr_1=2;

dtscltst_2=18846; kanker_2='Nee'; coecumwur_2=2; srtscr_2=1;
lnc_clnscp_2=12;
clean_2=1; scopie_2=9; npoliep_2=0; advice_2=99; ;
```

```
FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1074' THEN DO;
stoetnr=390029; dtscltst_1=18536; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=5; npoliep_1=0; advice_1=1; adviceyr_1=1;

dtscltst_2=18911; kanker_2='Nee'; coecumwur_2=2; srtscr_2=1;
lnc_clnscp_2=12;
clean_2=1; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1074' THEN DO;
stoetnr=390029; dtscltst_1=18536; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=5; npoliep_1=0; advice_1=1; adviceyr_1=1;

dtscltst_2=18911; kanker_2='Nee'; coecumwur_2=2; srtscr_2=1;
lnc_clnscp_2=12;
clean_2=1; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

* vanaf hier scopiedata checken!;
IF deelnr='A1079' THEN DO;
stoetnr=3030042; dtscltst_1=18030; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=5; npoliep_1=0; advice_1=1; adviceyr_1=1;

dtscltst_2=18205; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=5; npoliep_2=0; advice_2=1; adviceyr_2=2;

dtscltst_3=19088; kanker_3='Nee'; coecumwur_3=1; srtscr_3=1;
clean_3=3; scopie_3=5; npoliep_3=0; advice_3=3;

FORMAT dtscltst_1 dtscltst_2 dtscltst_3 date9. ;
END;

IF deelnr='A1089' THEN DO;
stoetnr=2890011; dtscltst_1=18322; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=7; npoliep_1=4; grpolsc1_1=3; grpolsc2_1=2;
grpolsc3_1=2; grpolsc4_1=2;
```

```
grpolpa1_1=99; grpolpa2_1=99; grpolpa3_1=99; grpolpa4_1=99; polpect_1=3;
PArap_1=2; date_PA_1=18323; PAnummer_poliep_1='T10-02211';
srtpoliep1_1=2; srtadenoma1_1=99; locp1_1=7; dysplasieform1_1=2;
srtpoliep2_1=1; locp2_1=3;
srtpoliep3_1=1; locp3_1=12;
srtpoliep4_1=1; locp4_1=4; advice_1=3;

dtscltst_2=18931; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date_PA_1 date9. ;
END;

IF deelnr='A1099' THEN DO;
stoetnr=2890010; dtscltst_1=17667; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=2; scopie_1=7; npoliep_1=4; grpolsc1_1=4; grpolsc2_1=4;
grpolsc3_1=4; grpolsc4_1=4;
grpolpa1_1=99; grpolpa2_1=99; grpolpa3_1=99; grpolpa4_1=99; polpect_1=3;
PArap_1=2; date_PA_1=17667; PAnummer_poliep_1='T08-04651';
srtpoliep1_1=2; srtadenoma1_1=99; locp1_1=7; dysplasieform1_1=2;
srtpoliep2_1=2; srtadenoma2_1=99; locp2_1=7; dysplasieform2_1=2;
srtpoliep3_1=2; srtadenoma3_1=99; locp3_1=8; dysplasieform3_1=2;
srtpoliep4_1=1; locp4_1=3; advice_1=3;

dtscltst_2=18323; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=2; scopie_2=7; npoliep_2=1; polpect_2=3; PArap_2=1; srtpoliep1_2=2;
srtadenoma1_2=1; locp1_2=5; dyplasieform1_2=2; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date_PA_1 date9. ;
END;

IF deelnr='A1123' THEN DO;
stoetnr=2680004; dtscltst_1=18234; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=99; grpolpa1_1=99;
polpect_1=3; PArap_1=1; date_PA_1=17667;
srtpoliep1_1=2; srtadenoma1_1=1; locp1_1=5; dysplasieform1_1=2;
advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 dtscltst_2 date_PA_1 date9. ;
END;

IF deelnr='A1133' THEN DO;
stoetnr=1060071; ovldt=17472;
FORMAT ovldt date9. ;
END;

IF deelnr='A1220' THEN DO;
```

```
stoetnr=1060028; dtscltst_1=18199; kanker_1='Nee'; coecumwur_1=2;
srtscr_1=1; lnc_clnscp_1=12;
clean_1=2; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1227' THEN DO;
stoetnr=1060033; dtscltst_1=18297; kanker_1='Nee'; coecumwur_1=2;
srtscr_1=1; lnc_clnscp_1=7;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=99; grpolpa1_1=99;
polpect_1=3; PArap_1=2; date_PA_1=18297; PAnummer_poliep_1='H10-5030';
srtpoliep1_1=2; srtadenoma1_1=1; locp1_1=12; dysplasieform1_1=2;
advice_1=1; adviceyr_1=2;

dtscltst_2=18323; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=1; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date_PA_1 date9. ;
END;

IF deelnr='A1138' THEN DO;
stoetnr=230051; dtscltst_1=18240; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=99; polpect_1=2; locp1_1=7;
advice_1=1; adviceyr_1=1;

dtscltst_2=18617; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1143' THEN DO;
stoetnr=1210043; dtscltst_1=18513; kanker_1='Nee'; coecumwur_1=2;
srtscr_1=1; lnc_clnscp_1=12;
clean_1=3; scopie_1=8; npoliep_1=0; advice_1=3;
FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1177' THEN DO;
stoetnr=1020049; dtscltst_1=18149; kanker_1='Nee'; coecumwur_1=9;
srtscr_1=99;
clean_1=3; scopie_1=8; npoliep_1=0; advice_1=3;

dtscltst_2=18563; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=1;

dtscltst_3=18970; kanker_3='Nee'; coecumwur_3=1; srtscr_3=1;
```

```
clean_3=3; scopie_3=8; npoliep_3=0; advice_3=1; adviceyr_3=1;

FORMAT dtscltst_1 dtscltst_2 dtscltst_3 date9. ;
END;

IF deelnr='A1188' THEN DO;
stoetnr=140040; dtscltst_1=18127; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=7; npoliep_1=3; grpolsc1_1=99; grpolsc2_1=99;
grpolsc3_1=99;
grpolpa1_1=99; grpolpa2_1=99; grpolpa3_1=99; polpect_1=3; PArap_1=1;
srtpoliep1_1=3; srtadenoma1_1=99; locp1_1=4; dysplasieform1_1=99;
srtpoliep2_1=3; srtadenoma2_1=99; locp2_1=4; dysplasieform2_1=99;
srtpoliep3_1=3; srtadenoma3_1=99; locp3_1=5; dysplasieform3_1=99;
advice_1=3;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1191' THEN DO;
stoetnr=1580004; dtscltst_1=18336; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=1;

dtscltst_2=18700; kanker_2='Nee'; coecumwur_2=2; srtscr_2=1;
lnc_clnscp_2=12;
clean_2=1; scopie_2=9; npoliep_2=0; advice_2=3;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1213' THEN DO;
stoetnr=860024; dtscltst_1=18063; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=2; scopie_1=7; npoliep_1=1; grpolsc1_1=99; grpolpa1_1=99;
polpect_1=3; PArap_1=2; date_PA_1=18063; PAnummer_poliep_1='T09-15614';
srtpoliep1_1=1; locp1_1=2; advice_1=1; adviceyr_1=2;

dtscltst_2=18744; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date_PA_1 date9. ;
END;

IF deelnr='A1214' THEN DO;
stoetnr=380027; dtscltst_1=18140; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=99; polpect_1=88;
srtpoliep1_1=1; locp1_1=4; advice_1=99;
```

```
FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1218' THEN DO;
stoetnr=2970054; dtscltst_1=18245; kanker_1='Nee'; coecumwur_1=2;
srtscr_1=1; lnc_clnscp_1=12;
clean_1=1; scopie_1=7; npoliep_1=2; grpolsc1_1=99; grpolsc2_1=99;
grpolpa1_1=99; grpolpa2_1=99; polpect_1=3; PArap_1=1;
srtpoliep1_1=1; locp1_1=12;
srtpoliep2_1=1; locp2_1=12;  advice_1=1; adviceyr_1=2;

dtscltst_2=18974; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=1; scopie_2=7; npoliep_2=3; grpolsc1_2=99; grpolsc2_2=99;
grpolsc3_2=99;
grpolpa1_2=99; grpolpa2_2=99; grpolpa3_2=99; polpect_2=3; PArap_2=1;
srtpoliep1_2=1; locp1_2=12; polcm1_2=15;
srtpoliep2_2=1; locp2_2=12; polcm2_2=15;
srtpoliep2_2=1; locp2_2=12; polcm3_2=15;
advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1219' THEN DO;
stoetnr=2970032; dtscltst_1=18225; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=8; npoliep_1=0; advice_1=3;

dtscltst_2=19003; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=7; npoliep_2=1; grpolsc1_2=10;  polpect_2=88;
srtpoliep1_2=2; srtadenoma1_2=99; locp1_2=12; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1223' THEN DO;
stoetnr=240017; dtscltst_1=18225; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

dtscltst_2=18940; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=1; scopie_2=7; npoliep_2=2; grpolsc1_2=99; grpolsc2_2=99;
grpolpa1_2=99; grpolpa2_2=99; polpect_2=3; PArap_2=2; date_PA_2=18940;
PAnummer_poliep_2='T11-10525';
srtpoliep1_2=2; srtadenoma1_2=99; locp1_2=7;
srtpoliep2_2=2; srtadenoma2_2=99; locp2_2=4;
advice_2=3;
```

```
FORMAT dtscltst_1 dtscltst_2 date_PA_2 date9. ;
END;

IF deelnr='A1233' THEN DO;
stoetnr=770060; dtscltst_1=18520; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=99; grpolpa1_1=99;
polpect_1=3; PArap_1=2; date_PA_1=18520; PAnummer_poliep_1='T10-06608';
srtpoliep1_1=1; locp1_1=5; advice_1=99;

FORMAT dtscltst_1 date_PA_1 date9. ;
END;

IF deelnr='A1240' THEN DO;
stoetnr=340023; dtscltst_1=18156; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

dtscltst_2=18702; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=1; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1241' THEN DO;
stoetnr=3090011; dtscltst_1=18159; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=4; grpolpa1_1=99;
polpect_1=3; PArap_1=1;
srtpoliep1_1=2; srtadenoma1_1=99; polcm1_1=10; advice_1=1; adviceyr_1=1;

dtscltst_2=18385; kanker_2='Nee'; coecumwur_2=2; srtscr_2=1;
lnc_clnscp_2=12;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=1;

dtscltst_3=18792; kanker_3='Nee'; coecumwur_3=2; srtscr_3=1;
lnc_clnscp_3=99;
clean_3=1; scopie_3=8; npoliep_3=0; advice_3=99;

FORMAT dtscltst_1 dtscltst_2 dtscltst_3 date9. ;
END;

IF deelnr='A1245' THEN DO;
stoetnr=940035; dtscltst_1=18298; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=7; npoliep_1=1; grpolsc1_1=99; grpolpa1_1=99;
polpect_1=3; PArap_1=2;
srtpoliep1_1=1; locp1_1=5;  advice_1=99;
```

```
FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1267' THEN DO;
stoetnr=770044; dtscltst_1=18241; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=7; npoliep_1=3; grpolsc1_1=99; grpolsc2_1=99;
grpolsc3_1=99; grpolpa1_1=99; grpolpa2_1=99; grpolpa3_1=99; polpect_1=3;
PArap_1=1; date_PA_1=18011; PAnummer_poliep_1='T09-5789';
srtpoliep1_1=1; srtpoliep2_1=1; polcm1_1=30; polcm2_1=15; locp1_1=4;
locp2_1=12; advice_1=99;

FORMAT dtscltst_1 date_PA_1 date9. ;
END;

IF deelnr='A1275' THEN DO;
stoetnr=1110015; dtscltst_1=18009; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=7; npoliep_1=2; grpolsc1_1=99; grpolsc2_1=99;
grpolpa1_1=99; grpolpa2_1=99; polpect_1=3; PArap_1=2;
polcm1_1=16; locp1_1=5; locp2_1=5; advice_1=99;

dtscltst_2=18893; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=2; scopie_2=7; npoliep_2=1; grpolsc1_2=8; grpolpa1_2=99;
polpect_2=3; parap_2=1;
srtpoliep1_2=2; srtadenoma1_2=99; locp1_2=5; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1280' THEN DO;
stoetnr=610140; dtscltst_1=18234; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=99; grpolpa1_1=99;
polpect_1=3; PArap_1=2;
srtpoliep1_1=1; polcm1_1=20; advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1311' THEN DO;
stoetnr=1770029; dtscltst_1=17924; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=5; npoliep_1=0; advice_1=99;

dtscltst_2=18714; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=1; scopie_2=7; npoliep_2=1; grpolsc1_2=4; grpolpa1_2=99;
polpect_2=3; PArap_2=2; date_PA_2=18716; PAnummer_poliep_2='T11-08833';
srtpoliep1_2=1; polcm1_2=70; locp1_2=2;
```

```
advice_2=99;

FORMAT dtscltst_1 dtscltst_2 date_PA_2 date9. ;
END;

IF deelnr='A1311' THEN DO;
stoetnr=2580008;
END;

IF deelnr='A1357' THEN DO;
stoetnr=2580008; dtscltst_1=18248; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=3; grpolpa1_1=99;
polpect_1=3; PArap_1=2;
srtpoliep1_1=2; srtadenoma1_1=1; locp1_1=5; advice_1=1; adviceyr_1=1;

dtscltst_2=18613; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=1;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1365' THEN DO;
stoetnr=690028;
END;

IF deelnr='A1389' THEN DO;
stoetnr=70089; uterus=1; adnex=1; date_uterus=17532;

dtscltst_1=18134; kanker_1='Nee'; coecumwur_1=1; srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=3; grpolpa1_1=99;
polpect_1=88;
srtpoliep1_1=2; srtadenoma1_1=99; locp1_1=4; dysplasieform1_1=2;
advice_1=1; adviceyr_1=2;

dtscltst_2=18841; kanker_2='Nee'; coecumwur_2=2; srtscr_2=1;
lnc_clnscp_2=1;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT date_uterus dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1435' THEN DO;
stoetnr=70058; dtscltst_1=18246; kanker_1='Nee'; coecumwur_1=1; srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

dtscltst_2=18962; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=2; scopie_2=7; npoliep_2=1; grpolsc1_2=4; grpolpa1_2=3;
polpect_2=3; PArap_2=2; date_PA_2=18963; PAnummer_poliep_2='T11-54753';
```

```
srtpoliep1_2=2; srtadenoma1_2=1;  locp1_2=7; dysplasieform1_2=2;
advice_2=1; adviceyr_2=1;

FORMAT dtscltst_1 dtscltst_2 date_PA_2 date9. ;
END;

IF deelnr='A1436' THEN DO;
stoetnr=70059; dtscltst_1=18232; kanker_1='Nee'; coecumwur_1=2; srtscr_1=1;
lnc_clnscp_1=99;
clean_1=2; scopie_1=7; npoliep_1=1; grpolsc1_1=99; grpolpa1_1=99;
polpect_1=3; PArap_1=2; date_PA_1=18233; PAnummer_poliep_1='T09-49934';
srtpoliep1_1=2; srtadenoma1_1=1; polcm1_1=40; locp1_1=4;
dysplasieform1_1=2; advice_1=99;

dtscltst_2=18281; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
lnc_clnscp_2=8;
clean_2=1; scopie_2=7; npoliep_2=1; grpolsc1_2=8; grpolpa1_2=99;
polpect_2=3; parap_2=1;
srtpoliep1_2=7; polcm1_2=30; locp1_2=4; advice_2=99;

FORMAT dtscltst_1 dtscltst_2 date_PA_1 date9. ;
END;

IF deelnr='A1422' THEN DO;
stoetnr=540020; dtscltst_1=18001; kanker_1='Nee'; coecumwur_1=9;
srtscr_1=99;
clean_1=3; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

dtscltst_2=18688; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=8; npoliep_2=0; advice_2=1; adviceyr_2=2;

FORMAT dtscltst_1 dtscltst_2 date9. ;
END;

IF deelnr='A1423' THEN DO;
stoetnr=740017; dtscltst_1=18294; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1473' THEN DO;
stoetnr=3840009; dtscltst_1=18316; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;
```

```
* A1437, A2118 en A2120 hebben poliep/kanker voor STARTvariabele. Alles wat
voor START variabele ligt, telt niet mee,
daarom handmatig gegevens verwijdert, zodat deze personen niet als case
worden meegerekend maar tijd tot aan
ovl (A1437), tijd tot aan 4e scopie (A2118) en tijd tot aan 2e scopie
(A2120) wordt uitgerekend.;
IF deelnr='A1437' THEN DO;
stoetnr=70067; dtscltst_1=17559; kanker_1='Nee'; ICD_1='.'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A2118' THEN DO;
stoetnr=3170021; dtscltst_1=17218; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0;
srtpoliep1_1='.'; srtpoliep2_1='.'; srtpoliep3_1='.'; srtpoliep4_1='.';
advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A2120' THEN DO;
stoetnr=3880012; dtscltst_1=17497; kanker_1='Nee'; ICD_1='.';
coecumwur_1=1; srtscr_1=1;
clean_1=1; scopie_1=8; npoliep_1=0;
srtpoliep1_1='.'; srtpoliep2_1='.'; srtpoliep3_1='.'; srtpoliep4_1='.';
advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1476' THEN DO;
stoetnr=4030034; dtscltst_1=18260; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=8; npoliep_1=0; advice_1=1; adviceyr_1=2;

FORMAT dtscltst_1 date9. ;
END;

IF deelnr='A1499' THEN DO;
stoetnr=3450008; dtscltst_1=18947; kanker_1='Nee'; coecumwur_1=1;
srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=2; grpolpa1_1=99;
polpect_1=3; PArap_1=2; date_PA_1=18950; PAnummer_poliep_1='T11-21054';
```

```
srtpoliep1_1=2; srtadenoma1_1=1; polcm1_1=15; locp1_1=4;
dysplasieform1_1=2; advice_1=1; adviceyr_1=1;

FORMAT dtscltst_1 date_PA_1 date9. ;
END;

IF deelnr='A2095' THEN DO;
stoetnr=3760019;
END;

IF deelnr='A1153' THEN DO;
stoetnr=3220059;
END;

IF deelnr='A1135' THEN DO;
stoetnr=870021;
END;

IF deelnr='A1391' THEN DO;
stoetnr=70066; dtscltst_1=17597; kanker_1='Nee'; coecumwur_1=1; srtscr_1=1;
clean_1=3; scopie_1=7; npoliep_1=1; grpolsc1_1=2; grpolpa1_1=99;
polpect_1=3; PArap_1=2; date_PA_1=17597; PAnummer_poliep_1='T08-9766';
srtpoliep1_1=2; srtadenoma1_1=1; polcm1_1=50; locp1_1=7;
dysplasieform1_1=99; advice_1=1; adviceyr_1=2;

dtscltst_2=18361; kanker_2='Nee'; coecumwur_2=1; srtscr_2=1;
clean_2=3; scopie_2=6; npoliep_2=0; advice_2=1; adviceyr_2=1;

dtscltst_3=18741; kanker_3='Nee'; coecumwur_3=1; srtscr_3=1;
clean_3=3; scopie_3=8; npoliep_3=0; advice_3=1; adviceyr_3=1;

FORMAT dtscltst_1 dtscltst_2 dtscltst_3 date_PA_1 date9. ;
END;
RUN;
* GELUKT;

* Checken of het uitmaakt welke variabele (kanker=ja of ICD code variabele)
je gebruikt voor aantal kankergevallen.
Voor uitkomst, zie ook 'puntjes analyse2.doc';
PROC PRINT DATA=total5;
VAR kanker_1 ICD_1 kanker_2 ICD_2 kanker_3 ICD_3 kanker_4 ICD_4 kanker_5
ICD_5 kanker_6 ICD_6
kanker_7 ICD_7 kanker_8 ICD_8;
RUN;

PROC FREQ DATA=total5;
TABLES kanker_1 ICD_1 kanker_2 ICD_2 kanker_3 ICD_3 kanker_4 ICD_4 kanker_5
ICD_5 kanker_6 ICD_6
kanker_7 ICD_7 kanker_8 ICD_8;
```

```
RUN;

* Aanmaken nieuwe variabelen voor adenomatous polyp, cancer and death cases
FU
489 observations;
DATA total5_fu;
SET total55;

IF srtpoliep1_1 IN(2,3,4,5,6) OR srtpoliep2_1 IN(2,3,4,5,6) OR srtpoliep3_1
IN(2,3,4,5,6) OR
srtpoliep4_1 IN(2,3,4,5,6) OR srtpoliep5_1 IN(2,3,4,5,6) OR srtpoliep6_1
IN(2,3,4,5,6) THEN CRACaseFU1=1;
ELSE CRACaseFU1=0;

IF srtpoliep1_2 IN(2,3,4,5,6) OR srtpoliep2_2 IN(2,3,4,5,6) OR srtpoliep3_2
IN(2,3,4,5,6) OR
srtpoliep4_2 IN(2,3,4,5,6) OR srtpoliep5_2 IN(2,3,4,5,6) OR srtpoliep6_2
IN(2,3,4,5,6) THEN CRACaseFU2=1;
ELSE CRACaseFU2=0;

IF srtpoliep1_3 IN(2,3,4,5,6) OR srtpoliep2_3 IN(2,3,4,5,6) OR srtpoliep3_3
IN(2,3,4,5,6) OR
srtpoliep4_3 IN(2,3,4,5,6) OR srtpoliep5_3 IN(2,3,4,5,6) OR srtpoliep6_3
IN(2,3,4,5,6) THEN CRACaseFU3=1;
ELSE CRACaseFU3=0;

IF srtpoliep1_4 IN(2,3,4,5,6) OR srtpoliep2_4 IN(2,3,4,5,6) OR srtpoliep3_4
IN(2,3,4,5,6) OR
srtpoliep4_4 IN(2,3,4,5,6) OR srtpoliep5_4 IN(2,3,4,5,6) OR srtpoliep6_4
IN(2,3,4,5,6) THEN CRACaseFU4=1;
ELSE CRACaseFU4=0;

IF srtpoliep1_5 IN(2,3,4,5,6) OR srtpoliep2_5 IN(2,3,4,5,6) OR srtpoliep3_5
IN(2,3,4,5,6) OR
srtpoliep4_5 IN(2,3,4,5,6) OR srtpoliep5_5 IN(2,3,4,5,6) OR srtpoliep6_5
IN(2,3,4,5,6) THEN CRACaseFU5=1;
ELSE CRACaseFU5=0;

IF srtpoliep1_6 IN(2,3,4,5,6) OR srtpoliep2_6 IN(2,3,4,5,6) OR srtpoliep3_6
IN(2,3,4,5,6) OR
srtpoliep4_6 IN(2,3,4,5,6) OR srtpoliep5_6 IN(2,3,4,5,6) OR srtpoliep6_6
IN(2,3,4,5,6) THEN CRACaseFU6=1;
ELSE CRACaseFU6=0;

IF srtpoliep1_7 IN(2,3,4,5,6) OR srtpoliep2_7 IN(2,3,4,5,6) OR srtpoliep3_7
IN(2,3,4,5,6) OR
srtpoliep4_7 IN(2,3,4,5,6) OR srtpoliep5_7 IN(2,3,4,5,6) OR srtpoliep6_7
IN(2,3,4,5,6) THEN CRACaseFU7=1;
ELSE CRACaseFU7=0;
```

```
IF srtpoliep1_8 IN(2,3,4,5,6) OR srtpoliep2_8 IN(2,3,4,5,6) OR srtpoliep3_8
IN(2,3,4,5,6) OR
srtpoliep4_8 IN(2,3,4,5,6) OR srtpoliep5_8 IN(2,3,4,5,6) OR srtpoliep6_8
IN(2,3,4,5,6) THEN CRACaseFU8=1;
ELSE CRACaseFU8=0;

* Gecheckt of alle missende waarden wel '.' toegekend hebben gekregen,
klopt;
IF ICD_1 NE '.' THEN CRCcaseFU1=1;
IF ICD_1 = '.' THEN CRCcaseFU1=0;
IF ICD_2 NE '.' THEN CRCcaseFU2=1;
IF ICD_2 = '.' THEN CRCcaseFU2=0;
IF ICD_3 NE '.' THEN CRCcaseFU3=1;
IF ICD_3 = '.' THEN CRCcaseFU3=0;
IF ICD_4 NE '.' THEN CRCcaseFU4=1;
IF ICD_4 = '.' THEN CRCcaseFU4=0;
IF ICD_5 NE '.' THEN CRCcaseFU5=1;
IF ICD_5 = '.' THEN CRCcaseFU5=0;
IF ICD_6 NE '.' THEN CRCcaseFU6=1;
IF ICD_6 = '.' THEN CRCcaseFU6=0;
IF ICD_7 NE '.' THEN CRCcaseFU7=1;
IF ICD_7 = '.' THEN CRCcaseFU7=0;
IF ICD_8 NE '.' THEN CRCcaseFU8=1;
IF ICD_8 = '.' THEN CRCcaseFU8=0;

IF ovl= 1 THEN CaseD=1;
ELSE CaseD=0;
RUN;

* A3006 is CRCcase3, maar datum diagnose is niet ingevuld. dtscltst
veranderen naar date diagnose;
DATA total5_fu1;
SET total5_fu;
IF deelnr='A3006' THEN datediag_3=dtscltst_3;
RUN;

* Aanpassen persontime voor persons met negatieve persontime. Kan hier wel
aanpassen in START variabele, want
seasonvariable is al eerder aangemaakt dus je verandert season niet.
Zie syntax7 voor hoe negatieve perstime gevonden.
Missende waardes voor personen die uiteindelijk (in latere dataset)
perstime 0 hadden. Nu in deze syntax
aangepast en syntax opnieuw gerunt. Ook dtscltst_1 op missing gezet, want
anders rekent ie alsnog perstime uit voor
deze personen, terwijl voor deze personen 1e scopie ligt voor START
variabele, dus niet meenemen;
DATA total5_fu1_1;
SET total5_fu1;
```

```
IF deelnr='A1437' THEN START=dtscltst_1;
IF deelnr='A2118' THEN START=dtscltst_1;
IF deelnr='A2120' THEN START=dtscltst_1;
IF deelnr='A1508' THEN START=dtscltst_1;
IF deelnr='A2133' THEN START=dtscltst_1;
IF deelnr='A2130' THEN START=dtscltst_1;

IF deelnr='A1487' THEN START='.';
IF deelnr='A1487' THEN dtscltst_1='.';

IF deelnr='A1159' THEN START='.';
IF deelnr='A1159' THEN dtscltst_1='.';

IF deelnr='A1507' THEN START=dtscltst_1;

IF deelnr='A1452' THEN START='.';
IF deelnr='A1452' THEN dtscltst_1='.';

IF deelnr='A1462' THEN START=dtscltst_1;
IF deelnr='A1488' THEN START=dtscltst_1;

IF deelnr='A1482' THEN START='.';
IF deelnr='A1482' THEN dtscltst_1='.';

IF deelnr='A1501' THEN START=dtscltst_1;
IF deelnr='A1470' THEN START=dtscltst_1;
IF deelnr='A2090' THEN START=dtscltst_1;
IF deelnr='A2135' THEN START=dtscltst_1;
IF deelnr='A2117' THEN START=dtscltst_1;
IF deelnr='A1458' THEN START=dtscltst_1;
RUN;

DATA total5_fu2;
SET total5_fu1_1;
IF CRACaseFU1=1 THEN perstime_fu_adenoom1=dtscltst_1-START;

IF CRACaseFU2=1 THEN perstime_fu_adenoom2=dtscltst_2-START;
IF CRACaseFU3=1 THEN perstime_fu_adenoom3=dtscltst_3-START;
IF CRACaseFU4=1 THEN perstime_fu_adenoom4=dtscltst_4-START;
IF CRACaseFU5=1 THEN perstime_fu_adenoom5=dtscltst_5-START;
IF CRACaseFU6=1 THEN perstime_fu_adenoom6=dtscltst_6-START;
IF CRACaseFU7=1 THEN perstime_fu_adenoom7=dtscltst_7-START;
IF CRACaseFU8=1 THEN perstime_fu_adenoom8=dtscltst_8-START;

IF CRCcaseFU1=1 THEN perstime_fu_cancer1=datediag_1-START;
IF CRCcaseFU2=1 THEN perstime_fu_cancer2=datediag_2-START;
IF CRCcaseFU3=1 THEN perstime_fu_cancer3=datediag_3-START;
IF CRCcaseFU4=1 THEN perstime_fu_cancer4=datediag_4-START;
IF CRCcaseFU5=1 THEN perstime_fu_cancer5=datediag_5-START;
```

```
IF CRCcaseFU6=1 THEN perstime_fu_cancer6=datediag_6-START;
IF CRCcaseFU7=1 THEN perstime_fu_cancer7=datediag_7-START;
IF CRCcaseFU8=1 THEN perstime_fu_cancer8=datediag_8-START;

IF ovl=1 THEN perstime_fu_death=ovldt-START;
RUN;

* Check of nu voor alle cases FU time is aangemaakt, klopt;
DATA total5_fu2_check;
SET total5_fu2;
KEEP deelnr CRACaseFU1 CRACaseFU2 CRACaseFU3 CRACaseFU4 CRACaseFU5
CRACaseFU6 CRACaseFU7 CRACaseFU8
CRCCaseFU1 CRCCaseFU2 CRCCaseFU3 CRCCaseFU4 CRCCaseFU5 CRCCaseFU6
CRCCaseFU7 CRCCaseFU8
perstime_fu_adenoom1 perstime_fu_adenoom2 perstime_fu_adenoom3
perstime_fu_adenoom4 perstime_fu_adenoom5
perstime_fu_adenoom6 perstime_fu_adenoom7 perstime_fu_adenoom8
perstime_fu_cancer1 perstime_fu_cancer2 perstime_fu_cancer3
perstime_fu_cancer4 perstime_fu_cancer5
perstime_fu_cancer6 perstime_fu_cancer7 perstime_fu_cancer8
perstime_fu_death START
dtscltst_1 dtscltst_2 dtscltst_3 dtscltst_4 dtscltst_5 dtscltst_6
dtscltst_7 dtscltst_8
datediag_1 datediag_2 datediag_3 datediag_4 datediag_5 datediag_6
datediag_7 datediag_8;
RUN;

* 489 observations;
DATA total5_fu3;
SET total5_fu2;
IF CRACaseFU1=1 OR CRACaseFU2=1 OR CRACaseFU3=1 OR CRACaseFU4=1 OR
CRACaseFU5=1 OR CRACaseFU6=1 OR CRACaseFU7=1 OR
CRACaseFU8=1 THEN CRAcaseFU=1;
ELSE CRAcaseFU=0;

IF CRCcaseFU1=1 OR CRCcaseFU2=1 OR CRCcaseFU3=1 OR CRCcaseFU4=1 OR
CRCcaseFU5=1 OR CRCcaseFU6=1 OR CRCcaseFU7=1 OR
CRCcaseFU8=1 THEN CRCcaseFU=1;
ELSE CRCcaseFU=0;
RUN;

PROC PRINT DATA=total5_fu3;
VAR deelnr CRAcaseFU CRCcaseFU START perstime_f;
RUN;

* Onderstaande datasets alleen voor aanmaken persontime en nr of scopies
tot aan 1e event! Deze datasets
(incl. variabele persontime en nr of scopies) samenvoegen en daarna
persontime en nr of scopie variabele invoegen in
```

```
total dataset (in total5_fu3).

* Voor aanmaken persontime adenome poliep cases
Deel A.B.
58 observations;
DATA total5_Acases1;
SET total5_fu3;
IF CRAcase=1;
RUN;

* Aantal scopien aanmaken
58 observations;
DATA total5_Acases1_sc;
SET total5_Acases1;
numbscopienew_e=numbscopienew_f+0;
RUN;

* Non adenome poliep cases deel 1 studie, voor deze personen kijken welke
FU adenome poliep OF
(als geen poliep gedurende FU period) cancer deel AB (CRTcase=1), cancer
deel FU, death, or no cases zijn
428 observations;
DATA total5_NoAcases1;
SET total5_fu3;
IF CRAcase=0;
RUN;


* Voor aanmaken persontime adenome poliep cases
Deel FU
66 observations;
DATA total5_Acases2;
SET total5_NoAcases1;
IF CRAcaseFU=1;
RUN;

DATA total5_Acases2a;
SET total5_Acases2;
DROP perstime_f;
RUN;

* Per scopie voor de cases van die scopie persontime en nr scopie tot aan
event aanmaken, want deze variabelen
alleen tot aan 1e event meenemen.
Daarna worden die cases eruit 'gefilterd', zodat alleen persontime en nr of
scopies tot aan scopie waarin
event is berekend wordt.
```

```
Totaal aantal van de losse adenome polyp cases moet gelijk zijn aan totaal
poliep cases (=66), klopt;

* Adenome polypcases 1e scopie
27 observations;
DATA total5_Acases2_1;
SET total5_Acases2a;
IF CRAcaseFU1=1;
RUN;

* Perstime aanmaken
27 observations;
DATA total5_Acases2_1_perstime;
SET total5_Acases2_1;
perstime_f=(perstime_fu_adenoom1)/(365.25/12);
RUN;

* Aantal scopien aanmaken
27 observations;
DATA total5_Acases2_1_perstime_sc;
SET total5_Acases2_1_perstime;
numbscopienew_e=numbscopienew_f+1;
RUN;

* 39 observations;
DATA total5_Acases2_1N;
SET total5_Acases2a;
IF CRAcaseFU1=0;
RUN;

* Adenome polypcases 2e scopie
* 32 observations;
DATA total5_Acases2_2;
SET total5_Acases2_1N;
IF CRAcaseFU2=1;
RUN;

* Perstime aanmaken
32 observations;
DATA total5_Acases2_2_perstime;
SET total5_Acases2_2;
perstime_f=(perstime_fu_adenoom2)/(365.25/12);
RUN;

* Aantal scopien aanmaken
32 observations;
DATA total5_Acases2_2_perstime_sc;
SET total5_Acases2_2_perstime;
numbscopienew_e=numbscopienew_f+2;
```

```
RUN;

* 7 observations;
DATA total5_Acases2_2N;
SET total5_Acases2_1N;
IF CRAcaseFU2=0;
RUN;

* Adenome polypcases 3e scopie
* 7 observations;
DATA total5_Acases2_3;
SET total5_Acases2_2N;
IF CRAcaseFU3=1;
RUN;

* Perstime aanmaken
7 observations;
DATA total5_Acases2_3_perstime;
SET total5_Acases2_3;
perstime_f=(perstime_fu_adenoom3)/(365.25/12);
RUN;

* Aantal scopien aanmaken
7 observations;
DATA total5_Acases2_3_perstime_sc;
SET total5_Acases2_3_perstime;
numbscopienew_e=numbscopienew_f+3;
RUN;

* 0 observations;
DATA total5_Acases2_3N;
SET total5_Acases2_2N;
IF CRAcaseFU3=0;
RUN;

* Non adenome poliep cases FU, voor deze personen kijken welke cancer cases
(deel AB of FU) zijn
362 observations;
DATA total5_NoAcases2;
SET total5_NoAcases1;
IF CRAcaseFU=0;
RUN;

* Voor aanmaken persontime cancer cases
Deel AB (CRTcase=1)
5 observations (klopt, want 65 CRT cases deel AB - 58 CRAcases deel AB,
maar van de 7 (65-58) CRC(?) cases deel AB zijn
2 CRAcaseFU dus tellen niet meer mee bij CRTcases);
DATA total5_Ccases1;
```

```
SET total5_NoAcases2;
IF CRTcase=1;
RUN;

* Aantal scopien aanmaken
5 observations;
DATA total5_Ccases1_sc;
SET total5_Ccases1;
numbscopienew_e=numbscopienew_f+0;
RUN;

* 357 observations;
DATA total5_NoCcases1;
SET total5_NoAcases2;
IF CRTcase=0;
RUN;

* Voor aanmaken persontime cancer cases
Deel FU
12 observations;
DATA total5_Ccases2;
SET total5_NoCcases1;
IF CRCcaseFU=1;
RUN;

DATA total5_Ccases2a;
SET total5_Ccases2;
DROP perstime_f;
RUN;

* Totaal aantal van de losse cancer cases moet gelijk zijn aan totaal
cancer cases (=12), klopt;

* Cancer cases 1e scopie
 7 observations;
DATA total5_Ccases2_1;
SET total5_Ccases2a;
IF CRCcaseFU1=1;
RUN;

* Perstime aanmaken
 7 observations;
DATA total5_Ccases2_1_perstime;
SET total5_Ccases2_1;
perstime_f=(perstime_fu_cancer1)/(365.25/12);
RUN;

* Aantal scopien aanmaken
7 observations;
```

```
DATA total5_Ccases2_1_perstime_sc;
SET total5_Ccases2_1_perstime;
numbscopienew_e=numbscopienew_f+1;
RUN;

* 5 observations;
DATA total5_Ccases2_1N;
SET total5_Ccases2a;
IF CRCcaseFU1=0;
RUN;

* Cancer cases 2e scopie
2 observations;
DATA total5_Ccases2_2;
SET total5_Ccases2_1N;
IF CRCcaseFU2=1;
RUN;

* Perstime aanmaken
2 observations;
DATA total5_Ccases2_2_perstime;
SET total5_Ccases2_2;
perstime_f=(perstime_fu_cancer2)/(365.25/12);
RUN;

* Aantal scopien aanmaken
2 observations;
DATA total5_Ccases2_2_perstime_sc;
SET total5_Ccases2_2_perstime;
numbscopienew_e=numbscopienew_f+2;
RUN;

* 3 observations;
DATA total5_Ccases2_2N;
SET total5_Ccases2_1N;
IF CRCcaseFU2=0;
RUN;

* Cancer cases 3e scopie
2 observations;
DATA total5_Ccases2_3;
SET total5_Ccases2_2N;
IF CRCcaseFU3=1;
RUN;

* Perstime aanmaken
2 observations;
DATA total5_Ccases2_3_perstime;
SET total5_Ccases2_3;
```

```
perstime_f=(perstime_fu_cancer3)/(365.25/12);
RUN;

* Aantal scopien aanmaken;
DATA total5_Ccases2_3_perstime_sc;
SET total5_Ccases2_3_perstime;
numbscopienew_e=numbscopienew_f+3;
RUN;

* 1 observations;
DATA total5_Ccases2_3N;
SET total5_Ccases2_2N;
IF CRCcaseFU3=0;
RUN;

* Cancer cases 4e scopie
1 observations;
DATA total5_Ccases2_4;
SET total5_Ccases2_3N;
IF CRCcaseFU4=1;
RUN;

* Perstime aanmaken
1 observations;
DATA total5_Ccases2_4_perstime;
SET total5_Ccases2_4;
perstime_f=(perstime_fu_cancer4)/(365.25/12);
RUN;

* Aantal scopien aanmaken
1 observations;
DATA total5_Ccases2_4_perstime_sc;
SET total5_Ccases2_4_perstime;
numbscopienew_e=numbscopienew_f+4;
RUN;

* 0 observations;
DATA total5_Ccases2_4N;
SET total5_Ccases2_3N;
IF CRCcaseFU4=0;
RUN;

* Non cancer cases (deel AB en FU), voor deze personen kijken welke death
cases zijn
345 observations;
DATA total5_NoCcases2;
SET total5_NoCcases1;
IF CRCcaseFU=0;
RUN;
```

```
* Voor aanmaken persontime death cases
Deel FU
7 observations;
DATA total5_Dcases2;
SET total5_NoCcases2;
IF CaseD=1;
RUN;

DATA total5_Dcases2a;
SET total5_Dcases2;
DROP perstime_f;
RUN;

* 7 observations;
DATA total5_Dcases2_perstime;
SET total5_Dcases2a;
perstime_f=(perstime_fu_death)/(365.25/12);
RUN;

* Aantal scopien aanmaken. Als +0, dan overleden voor datum 1e scopie
7 observations;
DATA total5_Dcases2_perstime_sc;
SET total5_Dcases2_perstime;
IF deelnr='A1028' THEN numbscopienew_e=numbscopienew_f+1;
IF deelnr='A1071' THEN numbscopienew_e=numbscopienew_f+0;
IF deelnr='A1130' THEN numbscopienew_e=numbscopienew_f+0;
IF deelnr='A1256' THEN numbscopienew_e=numbscopienew_f+0;
IF deelnr='A1437' THEN numbscopienew_e=numbscopienew_f+1;
IF deelnr='A3024' THEN numbscopienew_e=numbscopienew_f+0;
IF deelnr='A3026' THEN numbscopienew_e=numbscopienew_f+0;
RUN;

* Noncases FU, voor deze personen persontime tot aan laatste scopie
uitrekenen
338 observations;
DATA total5_noncases2;
SET total5_NoCcases2;
IF CaseD=0;
RUN;

* Om te checken vanaf welke scopie terugtellen. Laatste persoon noncase
uiterlijk 4 scopieen gehad;
DATA total5_noncases2_check;
SET total5_noncases2;
KEEP deelnr dtscltst_1 dtscltst_2 dtscltst_3 dtscltst_4 dtscltst_5
dtscltst_6 dtscltst_7 dtscltst_8;
RUN;
```

```
* Aanmaken perstime_f voor noncases
Personen 4e scopie
338 observations;
DATA Total5_noncases_sc4;
SET Total5_noncases2;
IF dtscltst_4='.' THEN scopie4=0;
IF dtscltst_4 NE '.' THEN scopie4=1;
RUN;

* Personen die scopie 4 gehaald hebben 'eruit' geselecteerd
10 observations;
DATA Total5_noncases_sc4_1;
SET Total5_noncases_sc4;
IF scopie4=1;
RUN;

* Perstime aanmaken
10 observations;
DATA total5_noncases_sc4_1a;
SET total5_noncases_sc4_1;
DROP perstime_f;
RUN;

DATA total5_noncases_sc4_1_perstime;
SET total5_noncases_sc4_1a;
perstime_f=(dtscltst_4-START)/(365.25/12);
RUN;

* Aantal scopien aanmaken
10 observations;
DATA total5_noncases_sc4_1_pt_sc;
SET total5_noncases_sc4_1_perstime;
numbscopienew_e=numbscopienew_f+4;
RUN;

* Hiermee verder naar scopie 3, dit is de groep die niet tot aan scopie 4
gekomen is.
Terugtellen naar scopie 1
328 observations;
DATA Total5_noncases_sc4_0;
SET Total5_noncases_sc4;
IF scopie4=0;
RUN;

* Personen 3e scopie
328 observations;
DATA Total5_noncases_sc3;
SET Total5_noncases_sc4_0;
IF dtscltst_3='.' THEN scopie3=0;
```

```
IF dtscltst_3 NE '.' THEN scopie3=1;
RUN;

* Personen die scopie 3 gehaald hebben 'eruit' geselecteerd
34 observations;
DATA Total5_noncases_sc3_1;
SET Total5_noncases_sc3;
IF scopie3=1;
RUN;

* Perstime aanmaken
34 observations;
DATA total5_noncases_sc3_1a;
SET total5_noncases_sc3_1;
DROP perstime_f;
RUN;

DATA total5_noncases_sc3_1_perstime;
SET total5_noncases_sc3_1a;
perstime_f=(dtscltst_3-START)/(365.25/12);
RUN;

* Aantal scopien aanmaken
34 observations;
DATA total5_noncases_sc3_1_pt_sc;
SET total5_noncases_sc3_1_perstime;
numbscopienew_e=numbscopienew_f+3;
RUN;

* Hiermee verder naar scopie 2, dit is de groep die niet tot aan scopie 3
gekomen is.
Terugtellen naar scopie 1
294 observations;
DATA Total5_noncases_sc3_0;
SET Total5_noncases_sc3;
IF scopie3=0;
RUN;

* Personen 2e scopie
294 observations;
DATA Total5_noncases_sc2;
SET Total5_noncases_sc3_0;
IF dtscltst_2='.' THEN scopie2=0;
IF dtscltst_2 NE '.' THEN scopie2=1;
RUN;

* Personen die scopie 2 gehaald hebben 'eruit' geselecteerd
141 observations;
DATA Total5_noncases_sc2_1;
```

```
SET Total5_noncases_sc2;
IF scopie2=1;
RUN;

* Perstime aanmaken
141 observations;
DATA total5_noncases_sc2_1a;
SET total5_noncases_sc2_1;
DROP perstime_f;
RUN;

DATA total5_noncases_sc2_1_perstime;
SET total5_noncases_sc2_1a;
perstime_f=(dtscltst_2-START)/(365.25/12);
RUN;

* Aantal scopien aanmaken
141 observations;
DATA total5_noncases_sc2_1_pt_sc;
SET total5_noncases_sc2_1_perstime;
numbscopienew_e=numbscopienew_f+2;
RUN;

* Hiermee verder naar scopie 1, dit is de groep die niet tot aan scopie 2
gekomen is.
Terugtellen naar scopie 1
153 observations;
DATA Total5_noncases_sc2_0;
SET Total5_noncases_sc2;
IF scopie2=0;
RUN;

* Personen 1e scopie
153 observations;
DATA Total5_noncases_sc1;
SET Total5_noncases_sc2_0;
IF dtscltst_1='.' THEN scopie1=0;
IF dtscltst_1 NE '.' THEN scopie1=1;
RUN;

* Personen die scopie 1 gehaald hebben 'eruit' geselecteerd
141 observations;
DATA Total5_noncases_sc1_1;
SET Total5_noncases_sc1;
IF scopie1=1;
RUN;

* Perstime aanmaken
141 observations;
```

```
DATA total5_noncases_sc1_1a;
SET total5_noncases_sc1_1;
DROP perstime_f;
RUN;

DATA total5_noncases_sc1_1_perstime;
SET total5_noncases_sc1_1a;
perstime_f=(dtscltst_1-START)/(365.25/12);
RUN;

* Aantal scopien aanmaken
141 observations;
DATA total5_noncases_sc1_1_pt_sc;
SET total5_noncases_sc1_1_perstime;
numbscopienew_e=numbscopienew_f+1;
RUN;

* Hiermee verder naar scopie A.B., dit is de groep die niet tot aan scopie
1 FU gekomen is.
Voor deze groep de perstime_f uit 1e deel study meenemen
Noncases, geen FU scopie
12 observations;
DATA Total5_noncases_nofuscopie;
SET Total5_noncases_sc1;
IF scopie1=0;
RUN;

* Noncases, geen FU scopie, geen scopie deel A.B. Deze groep uit studie
laten
8 observations;
DATA total5_noncases_noscopie_atall;
SET total5_noncases_nofuscopie;
IF perstime_f = '.';
RUN;

* Noncases, geen FU scopie, wel scopie deel A.B. Perstime_f uit 1e deel
study nemen
4 observations;
DATA total5_noncases_scopiedeelAB;
SET total5_noncases_nofuscopie;
IF perstime_f NE '.';
RUN;

* Aantal scopien aanmaken
4 observations;
DATA total5_noncases_scopiedeelAB_sc;
SET total5_noncases_scopiedeelAB;
numbscopienew_e=numbscopienew_f+0;
RUN;
```

```
* Misschien deze syntax vergelijken met welke deelnemers Renate wel/geen
scopie heeft;
DATA total5_noncases_nofuscopie_check;
SET total5_noncases_nofuscopie;
KEEP deelnr perstime_f;
RUN;

* Combineren datasets met (nieuw aangemaakte) perstime_f
478 observations (486 - 8 (persons without scopie));
DATA compleet;
SET
total5_Acases1_sc
total5_Acases2_1_perstime_sc total5_Acases2_2_perstime_sc
total5_Acases2_3_perstime_sc
total5_Ccases1_sc
total5_Ccases2_1_perstime_sc total5_Ccases2_2_perstime_sc
total5_Ccases2_3_perstime_sc total5_Ccases2_4_perstime_sc
total5_Dcases2_perstime_sc
total5_noncases_scopiedeelAB_sc
total5_noncases_sc4_1_pt_sc total5_noncases_sc3_1_pt_sc
total5_noncases_sc2_1_pt_sc total5_noncases_sc1_1_pt_sc;
RUN;

* Samenvoegen nieuwe perstime_f in total dataset. Eerst perstime_f
verwijderen, want dit is oude perstime_f. Voor alle
personen nu nieuwe perstime_f aangemaakt (of oude gehouden: total5_Acases1,
total5_Ccases1 en
total5_noncases_scopiedeelAB), te vinden in dataset 'compleet'.
489 observations;
DATA total5_fu3_1;
SET total5_fu3;
DROP perstime_f;
RUN;

PROC SORT DATA=total5_fu3_1;
BY deelnr;
RUN;

PROC SORT DATA=compleet;
BY deelnr;
RUN;

DATA compleet_perstime;
MERGE total5_fu3_1 compleet;
BY deelnr;
RUN;

* Dataset met nieuw aangemaakte perstime_f. Hiermee verder gaan.
```

```
478 observations;
DATA compleet_perstime1;
SET compleet_perstime;
IF perstime_f NE '.';
RUN;

* De 11 personen zonder scopie. In deze dataset zelfde personen als in
total5_noncases_noscopie_atall (met 3 extra
personen, waarvoor alle invoer leeg is. Dit is de invoer die weggelaten is,
want eerst 489 personen in dataset;
DATA compleet_noperstime;
SET compleet_perstime;
IF perstime_f='.';
RUN;

* Aanmaken CRAcase variabele. Eerst splitsen deel A.B. en deel FU, omdat in
deel A.B. al CRAcase aangemaakt is
Deel A.B.
58 observations;
DATA compleet_perstime1_1;
SET compleet_perstime1;
IF CRAcase=1;
RUN;

* Deel FU
420 observations;
DATA compleet_perstime2;
SET compleet_perstime1;
IF CRAcase=0;
RUN;

DATA compleet_perstime2_1;
SET compleet_perstime2;
DROP CRAcase;
RUN;

DATA compleet_perstime2_2;
SET compleet_perstime2_1;
IF CRAcaseFU=1 THEN CRAcase=1;
IF CRAcaseFU=0 THEN CRAcase=0;
IF CRCcaseFU=1 THEN CRCcase=1;
IF CRCcaseFU=0 THEN CRCcase=0;
RUN;

PROC SORT DATA=compleet_perstime1_1;
BY deelnr;
RUN;

PROC SORT DATA=compleet_perstime2_2;
```

```
BY deelnr;
RUN;

* Dataset met nieuw aangemaakte perstime_f én cra en crc case variable.
Hiermee verder gaan.
478 observations;
DATA compleet_perstime_cracase;
SET compleet_perstime1_1 compleet_perstime2_2;
BY deelnr;
RUN;

* Overview supplement intake GeoLynch study A*B;
PROC FREQ DATA=compleet_perstime_cracase;
TABLES v85a * v85b / NOPERCENT NOCOL NOROW;
TABLES v87a * v87b / NOPERCENT NOCOL NOROW;
TABLES v88a * v88b / NOPERCENT NOCOL NOROW;
TABLES v89a * v89b / NOPERCENT NOCOL NOROW;
TABLES v90a * v90b / NOPERCENT NOCOL NOROW;
TABLES v91a * v91b / NOPERCENT NOCOL NOROW;
TABLES v92a * v92b / NOPERCENT NOCOL NOROW;
TABLES v93a * v93b / NOPERCENT NOCOL NOROW;
TABLES v94a * v94b / NOPERCENT NOCOL NOROW;
TABLES v95a * v95b / NOPERCENT NOCOL NOROW;
RUN;

* Overview supplement intake per supplement questionA;
PROC FREQ DATA=compleet_perstime_cracase;
TABLES v85a v86 v87a v88a v89a v90a v91a v92a v93a v94a v95a;
RUN;

* Overview supplement intake per supplement questionB;
PROC FREQ DATA=compleet_perstime_cracase;
TABLES v85b v87b v88b v89b v90b v91b v92b v93b v94b v95b;
RUN;

* Vervangen 2 foute invoer v87a en v94a;
DATA zoekv87a;
SET compleet_perstime_cracase;
IF v87a=0;
RUN;

DATA zoekv94a;
SET compleet_perstime_cracase;
IF v94a=0;
RUN;

DATA final;
SET compleet_perstime_cracase;
* Foute invoer supplementintake;
```

```
IF deelnr='A3031' THEN v87a=1;
IF deelnr='A1170' THEN v94a=1;
* Deze 6 mensen zijn overleden, en hebben geen vervolgscopie. Omdat anders
variabele no. of scopies
niet op 100% uitkwam, deze personen uit analyse gehaald (zie mail Renate
23-05)
A1437 heeft wel vervolgscopie, zie syntax hieronder 'PROC PRINT final1',
maar klopt eigenlijk niet.
Omdat datum diag kanker lag voor START var, START var verandert, en
handmatig A1437 noncase gemaakt. Perstime tot aan
overleden aangemaakt (zie mail Renate 08-05);
IF deelnr='A1071' THEN DELETE;
IF deelnr='A1130' THEN DELETE;
IF deelnr='A1133' THEN DELETE;
IF deelnr='A1437' THEN DELETE;
IF deelnr='A3024' THEN DELETE;
IF deelnr='A3026' THEN DELETE;
RUN;

* Overview supplement intake GeoLynch study A*B, check if it is now
correct;
PROC FREQ DATA=final;
TABLES v85a * v85b / NOPERCENT NOCOL NOROW;
TABLES v87a * v87b / NOPERCENT NOCOL NOROW;
TABLES v88a * v88b / NOPERCENT NOCOL NOROW;
TABLES v89a * v89b / NOPERCENT NOCOL NOROW;
TABLES v90a * v90b / NOPERCENT NOCOL NOROW;
TABLES v91a * v91b / NOPERCENT NOCOL NOROW;
TABLES v92a * v92b / NOPERCENT NOCOL NOROW;
TABLES v93a * v93b / NOPERCENT NOCOL NOROW;
TABLES v94a * v94b / NOPERCENT NOCOL NOROW;
TABLES v95a * v95b / NOPERCENT NOCOL NOROW;
RUN;

* Aanmaken categorie voor nr of scopie en nsaid use;
DATA final1;
SET final;
IF numbscopienew_e=0 THEN numbscopienew_e_cat=0;
IF numbscopienew_e=1 THEN numbscopienew_e_cat=1;
IF numbscopienew_e=2 THEN numbscopienew_e_cat=2;
IF numbscopienew_e GE 3 THEN numbscopienew_e_cat=3;

IF totalnsaid=0 THEN totalnsaid_cat=0;
IF totalnsaid=1 THEN totalnsaid_cat=0;
IF totalnsaid=9 THEN totalnsaid_cat='.';
IF totalnsaid IN(2,3,4) THEN totalnsaid_cat=1;
RUN;

PROC PRINT DATA=final1;
```

```
VAR deelnr ovl;
WHERE numbscopienew_e_cat=0;
RUN;

* Supplementuser any;
DATA final_any;
SET final1;
IF v85a GE 2 OR v87a GE 2 OR v88a GE 2 OR v89a GE 2 OR v90a GE 2 OR v91a GE
2 OR v92a GE 2 OR v93a GE 2 OR
v94a GE 2 OR v95a GE 2 THEN Suppluser=1;
ELSE Suppluser=0;

IF v85a GE 2 THEN v85aa=1;
ELSE v85aa=0;
IF v87a GE 2 THEN v87aa=1;
ELSE v87aa=0;
IF v88a GE 2 THEN v88aa=1;
ELSE v88aa=0;
IF v89a GE 2 THEN v89aa=1;
ELSE v89aa=0;
IF v90a GE 2 THEN v90aa=1;
ELSE v90aa=0;
IF v91a GE 2 THEN v91aa=1;
ELSE v91aa=0;
IF v92a GE 2 THEN v92aa=1;
ELSE v92aa=0;
IF v93a GE 2 THEN v93aa=1;
ELSE v93aa=0;
IF v94a GE 2 THEN v94aa=1;
ELSE v94aa=0;
IF v95a GE 2 THEN v95aa=1;
ELSE v95aa=0;
RUN;

* Check frequency suppluser (user=190, nonuser=282 totaal=472);
PROC FREQ DATA=final_any;
TABLES Suppluser;
RUN;

* Supplementuser vanaf 1x/wk;
DATA final_wk;
SET final1;
IF v85a IN(1,2,3) OR v87a IN(1,2,3) OR v88a IN(1,2,3) OR v89a IN(1,2,3) OR
v90a IN(1,2,3) OR v91a IN(1,2,3)
OR v92a IN(1,2,3) OR v93a IN(1,2,3) OR v94a IN(1,2,3) OR v95a IN(1,2,3)
THEN suppluser=0;
IF v85a='.' OR v87a='.' OR v88a='.' OR v89a='.' OR v90a='.' OR v91a='.' OR
v92a='.' OR v93a='.'
OR v94a='.' OR v95a='.' THEN suppluser=0;
```

```
IF v85a IN(4,5,6,7) OR v87a IN(4,5,6,7) OR v88a IN(4,5,6,7) OR v89a
IN(4,5,6,7) OR v90a IN(4,5,6,7)
OR v91a IN(4,5,6,7) OR v92a IN(4,5,6,7) OR v93a IN(4,5,6,7) OR v94a
IN(4,5,6,7) OR
v95a IN(4,5,6,7) THEN suppluser=1;

IF v85a='.' THEN v85aa=0;
IF v85a IN(1,2,3) THEN v85aa=0;
IF v85a IN(4,5,6,7) THEN v85aa=1;

IF v87a='.' THEN v87aa=0;
IF v87a IN(1,2,3) THEN v87aa=0;
IF v87a IN(4,5,6,7) THEN v87aa=1;

IF v88a='.' THEN v88aa=0;
IF v88a IN(1,2,3) THEN v88aa=0;
IF v88a IN(4,5,6,7) THEN v88aa=1;

IF v89a='.' THEN v89aa=0;
IF v89a IN(1,2,3) THEN v89aa=0;
IF v89a IN(4,5,6,7) THEN v89aa=1;

IF v90a='.' THEN v90aa=0;
IF v90a IN(1,2,3) THEN v90aa=0;
IF v90a IN(4,5,6,7) THEN v90aa=1;

IF v91a='.' THEN v91aa=0;
IF v91a IN(1,2,3) THEN v91aa=0;
IF v91a IN(4,5,6,7) THEN v91aa=1;

IF v92a='.' THEN v92aa=0;
IF v92a IN(1,2,3) THEN v92aa=0;
IF v92a IN(4,5,6,7) THEN v92aa=1;

IF v93a='.' THEN v93aa=0;
IF v93a IN(1,2,3) THEN v93aa=0;
IF v93a IN(4,5,6,7) THEN v93aa=1;

IF v94a='.' THEN v94aa=0;
IF v94a IN(1,2,3) THEN v94aa=0;
IF v94a IN(4,5,6,7) THEN v94aa=1;

IF v95a='.' THEN v95aa=0;
IF v95a IN(1,2,3) THEN v95aa=0;
IF v95a IN(4,5,6,7) THEN v95aa=1;
RUN;

* Check frequency suppluser (user=173, nonuser=299 totaal=472);
PROC FREQ DATA=final_wk;
```

```
TABLES Suppluser;
RUN;

* PROC SORT, want by statement gebruiken voor frequency table 1 and 2;
PROC SORT DATA=final_any;
BY CRAcase;
RUN;

PROC SORT DATA=final_wk;
BY CRAcase;
RUN;

PROC SORT DATA=final_any;
BY suppluser;
RUN;

PROC SORT DATA=final_wk;
BY suppluser;
RUN;

* Check aantallen CRA en CRCcases;
PROC FREQ DATA=final_any;
TABLES CRAcase CRCcase CaseD;
RUN;

* Het enige verschil in beide datasets is de manier van supplementuser
definieren.

1. Frequencies for table 1, supplementuser any;
PROC FREQ DATA=final_any;
TABLES sexe opleiding gen crchist crahist othcahist numbscopienew_e_cat
operatie dumoperatiepart dumoperatiesubtot
suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa v95aa
season dumphysacthigh
smoker totalnsaid_cat;
TITLE 'table1';
RUN;

* Aparte tabel maken om te testen voor menopausal status and postmenopausal
hormone use;
PROC FREQ DATA=final_any;
TABLES hormons menstcur;
WHERE sexe=2;
TITLE 'table1 women any suppl use';
RUN;

* 2.;
PROC FREQ DATA=final_any;
```

```
TABLES sexe opleiding gen crchist crahist othcahist numbscopienew_e_cat
operatie dumoperatiepart dumoperatiesubtot
suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa v95aa
season dumphysacthigh
smoker totalnsaid_cat;
BY CRAcase;
TITLE 'table1 by cracase';
RUN;

* Aparte tabel maken om te testen voor menopausal status and postmenopausal
hormone use;
PROC FREQ DATA=final_any;
TABLES hormons menstcur;
BY CRAcase;
WHERE sexe=2;
TITLE 'table1 women any suppl use by cracase';
RUN;

* 3. Frequencies for table 1, supplementuser vanaf 1x/wk, alleen voor
supplement intake,
de rest kan uit vorige syntax gehaald worden;
PROC FREQ DATA=final_wk;
TABLES suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa
v95aa;
RUN;

* 4.;
PROC FREQ DATA=final_wk;
TABLES suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa
v95aa;
BY CRAcase;
RUN;

* 5. Check distribution continuous variables;
PROC CHART DATA=final_any;
VBAR perstime_f AGE BMI length kcal SumOfvezel alltotveggies alltotfruit
redmeat alltotprocessedmeat SumOfalcohol;
RUN;

* 6. Check distribution continuous variables, compare mean with median,
different? than use median;
 PROC UNIVARIATE DATA=final_any;
 VAR perstime_f AGE BMI length kcal SumOfvezel alltotveggies alltotfruit
redmeat alltotprocessedmeat SumOfalcohol;
 TITLE 'table1 any suppl use continuous';
 RUN;

 * 7.;
 PROC UNIVARIATE DATA=final_any;
```

```
 VAR perstime_f AGE BMI length kcal SumOfvezel alltotveggies alltotfruit
redmeat alltotprocessedmeat SumOfalcohol;
 BY CRAcase;
 TITLE 'table1 any suppl use continuous by cracase';
 RUN;

* 22.;
PROC FREQ DATA=final_any;
TABLES sexe opleiding gen crchist crahist othcahist numbscopienew_e_cat
operatie dumoperatiepart dumoperatiesubtot
suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa v95aa
season dumphysacthigh
smoker totalnsaid_cat;
BY suppluser;
TITLE 'table2';
RUN;

* Aparte tabel maken om te testen voor menopausal status and postmenopausal
hormone use;
PROC FREQ DATA=final_any;
TABLES hormons menstcur;
BY suppluser;
WHERE sexe=2;
TITLE 'table2 women any suppl use';
RUN;

* 77. Check distribution continuous variables, compare mean with median,
different? than use median;
 PROC UNIVARIATE DATA=final_any;
 VAR perstime_f AGE BMI length kcal SumOfvezel alltotveggies alltotfruit
redmeat alltotprocessedmeat SumOfalcohol;
 BY suppluser;
 TITLE 'table2 any suppl use continuous';
 RUN;

 * 222.;
PROC FREQ DATA=final_wk;
TABLES sexe opleiding gen crchist crahist othcahist numbscopienew_e_cat
operatie dumoperatiepart dumoperatiesubtot
suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa v95aa
season dumphysacthigh
smoker totalnsaid_cat hormons menstcur;
BY suppluser;
RUN;

* 777. Check distribution continuous variables, compare mean with median,
different? than use median;
 PROC UNIVARIATE DATA=final_wk;
```

```
 VAR perstime_f AGE BMI length kcal SumOfvezel alltotveggies alltotfruit
redmeat alltotprocessedmeat SumOfalcohol;
 BY suppluser;
 RUN;

* Check average height women and men apart;
PROC UNIVARIATE DATA=final_any;
VAR length;
WHERE sexe=2;
TITLE 'length women';
RUN;

PROC UNIVARIATE DATA=final_any;
VAR length;
WHERE sexe=1;
TITLE 'length man';
RUN;

* Supplementintake by season;
 PROC SORT DATA=final_any;
 BY season;
 RUN;

PROC FREQ DATA=final_any;
TABLES suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa
v95aa;
BY season;
RUN;

* Supplementintake by gender (voor vergelijking met VCP);
 PROC SORT DATA=final_any;
 BY sexe;
 RUN;

PROC FREQ DATA=final_any;
TABLES suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa
v95aa;
BY sexe;
RUN;

* Supplementintake by agecat and gender (voor vergelijking met VCP),
zie onder data analyse -> data analyse -> vraagjes, printscreens, etc.;
DATA final_any_agecat;
SET final_any;
IF AGE GE 19 AND AGE LT 31 THEN agecat=1;
ELSE IF AGE GE 31 AND AGE LT 51 THEN agecat=2;
ELSE IF AGE GE 51 AND AGE LT 70 THEN agecat=3;
ELSE IF AGE LT 19 AND AGE GE 70 THEN agecat='.';
RUN;
```

```
PROC SORT DATA=final_any_agecat;
BY agecat;
RUN;

PROC SORT DATA=final_any_agecat;
BY sexe;
RUN;

PROC FREQ DATA=final_any_agecat;
TABLES suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa
v95aa;
BY sexe;
WHERE agecat=1;
TITLE 'agecat1';
RUN;

PROC FREQ DATA=final_any_agecat;
TABLES suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa
v95aa;
BY sexe;
WHERE agecat=2;
TITLE 'agecat2';
RUN;

PROC FREQ DATA=final_any_agecat;
TABLES suppluser v85aa v87aa v88aa v89aa v90aa v91aa v92aa v93aa v94aa
v95aa;
BY sexe;
WHERE agecat=3;
TITLE 'agecat3';
RUN;

* Zijn er veel mensen die meerdere supplementen tegelijk nemen?;
DATA final_any_supplcount;
SET final_any;
suppltotaal=v85aa+v87aa+v88aa+v89aa+v90aa+v91aa+v92aa+v93aa+v94aa+v95aa;
RUN;

PROC FREQ DATA=final_any_supplcount;
TABLES suppltotaal;
RUN;

* Testen voor confounders: confounder-disease (adenomatous polyps).
Gebaseerd op tabel 1 en 2 en of het
logische confounders zijn (biological plausibility);

* Age at study entry;
PROC PHREG DATA=final_any covs(aggregate);
```

```
MODEL perstime_f*CRAcase(0)= age / RL;
TITLE 'age';
ID famnr;
RUN;

* Sex;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= sexe / RL;
TITLE 'sex';
ID famnr;
RUN;

* Educational level;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= opleiding / RL;
TITLE 'educational level';
ID famnr;
RUN;

* BMI;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= BMI / RL;
TITLE 'educational level';
ID famnr;
RUN;

* Physical activity;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= dumphysacthigh / RL;
TITLE 'educational level';
ID famnr;
RUN;

* History of CRC;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= crchist / RL;
TITLE 'history of crc';
ID famnr;
RUN;

* History of adenomatous polyps;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= crahist / RL;
TITLE 'history of adenomatous polyps';
ID famnr;
RUN;

* History of other cancers;
PROC PHREG DATA=final_any covs(aggregate);
```

```
MODEL perstime_f*CRAcase(0)= othcahist / RL;
TITLE 'history of other cancers';
ID famnr;
RUN;

* No. of colonoscopies;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= numbscopienew_e_cat / RL;
TITLE 'no. of colonoscopies';
ID famnr;
RUN;

* Colon surgery partial;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= dumoperatiepart / RL;
TITLE 'colon surgery partial';
ID famnr;
RUN;

* Colon surgery subtotal;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= dumoperatiesubtot / RL;
TITLE 'colon surgery subtotal';
ID famnr;
RUN;

* Fruit intake;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= alltotfruit / RL;
TITLE 'fruit intake';
ID famnr;
RUN;

* Smoking habits;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= smoker / RL;
TITLE 'smoking habits';
ID famnr;
RUN;

* Alcohol intake;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= SumOfalcohol / RL;
TITLE 'alcohol intake';
ID famnr;
RUN;

* NSAID use;
PROC PHREG DATA=final_any covs(aggregate);
```

```
MODEL perstime_f*CRAcase(0)= totalnsaid_cat / RL;
TITLE 'nsaid use';
ID famnr;
RUN;

* Testen in model: educational level, history of CRC, history of
adenomatous polyps, history of other cancers,
no. of colonoscopies, smoking, NSAID use. Model with suppluser age and sex
is crude model;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding crchist crahist
othcahist numbscopienew_e_cat smoker
totalnsaid_cat / RL
SELECTION=BACKWARD;
TITLE 'final model backward method to select confounders';
ID famnr;
RUN;

* Testen in model: educational level, history of CRC, history of
adenomatous polyps, history of other cancers,
no. of colonoscopies, smoking, NSAID use. Model with suppluser age and sex
is crude model;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding crchist crahist
othcahist numbscopienew_e_cat smoker
totalnsaid_cat/ RL
SELECTION=FORWARD;
TITLE 'final model forward method to select confounders';
ID famnr;
RUN;

* Testen in model: educational level, history of CRC, history of
adenomatous polyps, history of other cancers,
no. of colonoscopies, smoking, NSAID use. Model with suppluser age and sex
is crude model;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding crchist crahist
othcahist numbscopienew_e_cat smoker
totalnsaid_cat/ RL
SELECTION=STEPWISE;
TITLE 'final model stepwise method to select confounders';
ID famnr;
RUN;

* Bij alledrie komen CRAhist, numbscopie en smoker eruit als confounders;

* Testen in model: confounders een voor een toevoegen;
* Crude model;
PROC PHREG DATA=final_any covs(aggregate);
```

```
MODEL perstime_f*CRAcase(0)= suppluser age sexe / RL;
TITLE 'Crude model';
ID famnr;
RUN;

* Crude model with education as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding / RL;
TITLE 'Crude model + education';
ID famnr;
RUN;

* Crude model with crc hist as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe crchist / RL;
TITLE 'Crude model + crc hist';
ID famnr;
RUN;

* Crude model with cra hist as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe crahist / RL;
TITLE 'Crude model + cra hist';
ID famnr;
RUN;

* Crude model with other cancer hist as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe othcahist / RL;
TITLE 'Crude model + othcahist';
ID famnr;
RUN;

* Crude model with no of scopies as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe numbscopienew_e_cat / RL;
TITLE 'Crude model + numbscopienew_e';
ID famnr;
RUN;

* Crude model with colon surgery partial as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe dumoperatiepart / RL;
TITLE 'Crude model + colon surgery partial';
ID famnr;
RUN;

* Crude model with colon surgery subtotal as confounder;
PROC PHREG DATA=final_any covs(aggregate);
```

```
MODEL perstime_f*CRAcase(0)= suppluser age sexe dumoperatiesubtot / RL;
TITLE 'Crude model + colon surgery subtotal';
ID famnr;
RUN;

* Crude model with smoker as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker / RL;
TITLE 'Crude model + smoker';
ID famnr;
RUN;

* Crude model with NSAID use as confounder;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe totalnsaid_cat / RL;
TITLE 'Crude model + NSAID use';
ID famnr;
RUN;

* None of the confounders seemed to be a confounder;

* Als een voor een gedaan, dan lijkt niks een confounder. Maar met forward,
backward and stepwise zijn
CRAhist, numbscopie en smoker confounders (en suppluser, age and sex worden
eruit gegooid).
Als je fully adjusted model runt (age, sex, crahist, numbscopie en smoker),
verandert HR bijna niks (zie syntax
ergens hieronder). SAS niet teveel vertrouwen, een voor een handmatig
toevoegen is dan beter;

* Een voor een toevoegen is een handmatige vorm van forward selection. Ook
nog combinaties testen;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding crahist/ RL;
TITLE 'Crude model + opleiding en crahist';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding
numbscopienew_e_cat / RL;
TITLE 'Crude model + opleiding en no of colonoscopies';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding smoker / RL;
TITLE 'Crude model + opleiding en smoker';
ID famnr;
```

```
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe crahist smoker / RL;
TITLE 'Crude model + crahist en smoker';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
/ RL;
TITLE 'Crude model + smoker en no of colonoscopies';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe crahist numbscopienew_e_cat
/ RL;
TITLE 'Crude model + crahist en no of colonoscopies';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding crahist
numbscopienew_e_cat / RL;
TITLE 'Crude model + opleiding, crahist en no of colonoscopies';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding crahist
numbscopienew_e_cat smoker/ RL;
TITLE 'Crude model + opleiding, crahist, no of colonoscopies en smoker';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe opleiding crahist smoker/
RL;
TITLE 'Crude model + opleiding, crahist en smoker';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe crahist smoker
numbscopienew_e_cat/ RL;
TITLE 'Crude model + crahist, smoker en no of scopies';
ID famnr;
RUN;
```

```
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe numbscopienew_e_cat smoker
opleiding/ RL;
TITLE 'Crude model + numbscopienew_e';
ID famnr;
RUN;

* Verschil crude met age and sex AND crude met age, sexe and smoker;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe / RL;
TITLE 'Crude model exl smoking';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker / RL;
TITLE 'Crude model incl smoking';
ID famnr;
RUN;

* None of the combinations of variables seemed to confound the association
between supplement intake and
colorectal adenomas. Continue met testen voor EM using crude model;
* Crude model is incl age, sexe, smoker, numbscopienew_e_cat;

* Possible EM: fruit + vegetable intake, history of colorectal carcinomas,
gender, smoking;

* Possible EM: Fruit + vegatable. Eerst beide variabelen samenvoegen en
quartiles uitrekenen;
DATA final_any_fruitveg;
SET final_any;
fruitveg= alltotveggies + alltotfruit;
RUN;

PROC UNIVARIATE DATA=final_any_fruitveg;
VAR fruitveg;
RUN;

* Continuous variables, dus indelen in categories;
DATA final_any_fruitveg1;
SET final_any_fruitveg;
IF fruitveg LT 179.1913 THEN fruitvegcat=1;
IF fruitveg GE 179.1913 AND fruitveg LT 282.8496 THEN fruitvegcat=2;
IF fruitveg GE 282.8496 AND fruitveg LT 387.1308 THEN fruitvegcat=3;
IF fruitveg GE 387.1308 THEN fruitvegcat=4;
RUN;
```

```
* Zet nonuser en q1 op HR=1.0. Voor de overige groepen HR uitrekenen. Zie
volgende syntax.;
DATA final_any_fruitveg2;
SET final_any_fruitveg1;
IF fruitvegcat=2 AND suppluser=0 THEN fruitvegnonuse_q2=1;
ELSE fruitvegnonuse_q2=0;
IF fruitvegcat=3 AND suppluser=0 THEN fruitvegnonuse_q3=1;
ELSE fruitvegnonuse_q3=0;
IF fruitvegcat=4 AND suppluser=0 THEN fruitvegnonuse_q4=1;
ELSE fruitvegnonuse_q4=0;

IF fruitvegcat=1 AND suppluser=1 THEN fruitveguse_q1=1;
ELSE fruitveguse_q1=0;
IF fruitvegcat=2 AND suppluser=1 THEN fruitveguse_q2=1;
ELSE fruitveguse_q2=0;
IF fruitvegcat=3 AND suppluser=1 THEN fruitveguse_q3=1;
ELSE fruitveguse_q3=0;
IF fruitvegcat=4 AND suppluser=1 THEN fruitveguse_q4=1;
ELSE fruitveguse_q4=0;
RUN;

* Suppluser en exposure variable (fruitvegcat) niet in model toevoegen.
Alleen age, sex, smoking, numbscopienew_e_cat en nieuw aangemaakte
variabele toevoegen. Uit deze syntax HR halen
(HR die achter nieuwe term staat);
PROC PHREG DATA=final_any_fruitveg2 covs(aggregate);
MODEL perstime_f*CRAcase(0)= age sexe smoker numbscopienew_e_cat
fruitvegnonuse_q2 fruitvegnonuse_q3 fruitvegnonuse_q4
fruitveguse_q1 fruitveguse_q2 fruitveguse_q3 fruitveguse_q4 / RL;
TITLE 'EM by fruitveg?';
ID famnr;
RUN;

* Testen voor P voor interactie: interactieterm aanmaken. Check p waarde
achter interactieterm;
* Note, dit kan evt. ook in 1 stap, zie ana epi map SAS handleiding, aparte
data stap niet perse nodig;
DATA final_any_fruitveg3;
SET final_any_fruitveg2;
suppluser_grfruit=suppluser*fruitvegcat;
RUN;

PROC PHREG DATA=final_any_fruitveg3 covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
fruitvegcat suppluser_grfruit / RL;
TITLE 'EM by fruitveg?';
ID famnr;
RUN;
```

```
* Possible EM: history of colorectal carcinomas (crthist);
DATA final_any_crthist;
SET final_any;
IF crthist=1 AND suppluser=0 THEN crtnonuse=1;
ELSE crtnonuse=0;

IF crthist=0 AND suppluser=1 THEN nocrtuse=1;
ELSE nocrtuse=0;
IF crthist=1 AND suppluser=1 THEN crtuse=1;
ELSE crtuse=0;
RUN;

* Suppluser en exposure variable (crthist) niet in model toevoegen.
Alleen age, sex, smoking, numbscopienew_e_cat en nieuw aangemaakte
variabele toevoegen. Uit deze syntax HR halen
(HR die achter nieuwe term staat);
PROC PHREG DATA=final_any_crthist covs(aggregate);
MODEL perstime_f*CRAcase(0)= age sexe smoker numbscopienew_e_cat crtnonuse
nocrtuse crtuse / RL;
TITLE 'EM by crthist?';
ID famnr;
RUN;

* Testen voor P voor interactie: interactieterm aanmaken. Check p waarde
achter interactieterm;
DATA final_any_crthist1;
SET final_any_crthist;
suppluser_crthist=suppluser*crthist;
RUN;

PROC PHREG DATA=final_any_crthist1 covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
crthist suppluser_crthist / RL;
TITLE 'EM by crthist?';
ID famnr;
RUN;

* Possible EM: gender;
DATA final_any_gender;
SET final_any;
IF sexe=2 AND suppluser=0 THEN femalenonuse=1;
ELSE femalenonuse=0;

IF sexe=1 AND suppluser=1 THEN maleuser=1;
ELSE maleuser=0;
IF sexe=2 AND suppluser=1 THEN femaleuser=1;
ELSE femaleuser=0;
```

```
RUN;

* Suppluser en exposure variable (sexe) niet in model toevoegen.
Alleen age, smoking, numbscopienew_e_cat en nieuw aangemaakte variabele
toevoegen. Uit deze syntax HR halen
(HR die achter nieuwe term staat);
PROC PHREG DATA=final_any_gender covs(aggregate);
MODEL perstime_f*CRAcase(0)= age smoker numbscopienew_e_cat femalenonuse
maleuser femaleuser / RL;
TITLE 'EM by gender?';
ID famnr;
RUN;

* Testen voor P voor interactie: interactieterm aanmaken. Check p waarde
achter interactieterm;
DATA final_any_gender1;
SET final_any_gender;
suppluser_sexe=suppluser*sexe;
RUN;

PROC PHREG DATA=final_any_gender1 covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
suppluser_sexe / RL;
TITLE 'EM by gender?';
ID famnr;
RUN;

* Possible EM: smoking;
DATA final_any_smoking;
SET final_any;
IF smoker=2 AND suppluser=0 THEN formnonuse=1;
ELSE formnonuse=0;
IF smoker=1 AND suppluser=0 THEN currnonuse=1;
ELSE currnonuse=0;

IF smoker=3 AND suppluser=1 THEN nevuse=1;
ELSE nevuse=0;
IF smoker=2 AND suppluser=1 THEN formuse=1;
ELSE formuse=0;
IF smoker=1 AND suppluser=1 THEN curruse=1;
ELSE curruse=0;
RUN;

* Suppluser en exposure variable (smoker) niet in model toevoegen.
Alleen age, sex, numbscopienew_e_cat en nieuw aangemaakte variabele
toevoegen. Uit deze syntax HR halen
(HR die achter nieuwe term staat);
PROC PHREG DATA=final_any_smoking covs(aggregate);
```

```
MODEL perstime_f*CRAcase(0)= age sexe numbscopienew_e_cat formnonuse
currnonuse nevuse formuse curruse / RL;
TITLE 'EM by smoking?';
ID famnr;
RUN;

* Testen voor P voor interactie: interactieterm aanmaken. Check p waarde
achter interactieterm;
DATA final_any_smoking1;
SET final_any_smoking;
suppluser_smoker=suppluser*smoker;
RUN;

PROC PHREG DATA=final_any_smoking1 covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
suppluser_smoker / RL;
TITLE 'EM by smoking?';
ID famnr;
RUN;

* None of the possible EM seemed to be EM.;

* Voor overzicht aantal cases tabel;

* Fruit and vegetable intake;
PROC SORT DATA=final_any_fruitveg3;
BY suppluser;
RUN;

PROC FREQ DATA=final_any_fruitveg3;
TABLES CRAcase;
BY suppluser;
WHERE fruitvegcat=1;
TITLE 'fruitvegcat1';
RUN;

PROC FREQ DATA=final_any_fruitveg3;
TABLES CRAcase;
BY suppluser;
WHERE fruitvegcat=2;
TITLE 'fruitvegcat2';
RUN;

PROC FREQ DATA=final_any_fruitveg3;
TABLES CRAcase;
BY suppluser;
WHERE fruitvegcat=3;
TITLE 'fruitvegcat3';
RUN;
```

```
PROC FREQ DATA=final_any_fruitveg3;
TABLES CRAcase;
BY suppluser;
WHERE fruitvegcat=4;
TITLE 'fruitvegcat4';
RUN;

* History of colorectal carcinomas;
PROC SORT DATA=final_any_crthist1;
BY suppluser;
RUN;

PROC FREQ DATA=final_any_crthist1;
TABLES CRAcase;
BY suppluser;
WHERE crthist=0;
TITLE 'no history of colorectal neoplasm';
RUN;

PROC FREQ DATA=final_any_crthist1;
TABLES CRAcase;
BY suppluser;
WHERE crthist=1;
TITLE 'history of colorectal neoplams';
RUN;

* Gender;
PROC SORT DATA=final_any_gender1;
BY suppluser;
RUN;

PROC FREQ DATA=final_any_gender1;
TABLES CRAcase;
BY suppluser;
WHERE sexe=1;
TITLE 'man';
RUN;

PROC FREQ DATA=final_any_gender1;
TABLES CRAcase;
BY suppluser;
WHERE sexe=2;
TITLE 'woman';
RUN;

* Smoking;
PROC SORT DATA=final_any_smoking1;
BY suppluser;
```

```
RUN;

PROC FREQ DATA=final_any_smoking1;
TABLES CRAcase;
BY suppluser;
WHERE smoker=3;
TITLE 'never smoker';
RUN;

PROC FREQ DATA=final_any_smoking1;
TABLES CRAcase;
BY suppluser;
WHERE smoker=2;
TITLE 'former smoker';
RUN;

PROC FREQ DATA=final_any_smoking1;
TABLES CRAcase;
BY suppluser;
WHERE smoker=1;
TITLE 'current smoker';
RUN;

* Testen HR voor vrouwen apart, zijn menopausal status and menopausal
hormone use confounders? Nee;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age smoker numbscopienew_e_cat / RL;
TITLE 'Crude model women any use';
ID famnr;
WHERE sexe=2;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age smoker numbscopienew_e_cat
hormons menstcur / RL;
TITLE 'Crude model women any use additionnaly adjusted for hormons and
menscur';
ID famnr;
WHERE sexe=2;
RUN;

PROC FREQ DATA=final_any;
TABLES suppluser;
WHERE sexe=2;
RUN;

PROC SORT DATA=final_any;
BY CRAcase;
RUN;
```

```
PROC FREQ DATA=final_any;
TABLES suppluser;
BY CRAcase;
WHERE sexe=2;
TITLE 'any supplement';
RUN;

*************************************************************************
************************************
*************************************************************************
************************************
****************************************** FINAL TEST
*********************************************************
*************************************************************************
************************************
*************************************************************************
************************************;

* Run crude model, also for multivit and vit c. Hoeft voor multivit and vit
c niet opnieuw voor confounders te testen,
want je wil modellen onderling ook kunnen vergelijken;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
/ RL;
TITLE 'Crude model any use';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= v85aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model multivit';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= v87aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model vit C';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= v92aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model calcium';
ID famnr;
RUN;
```

```
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= v94aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model visolie';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= v95aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model other';
ID famnr;
RUN;

* Fully adjusted models (gebaseerd op confounders by backward selection
method, om te vergelijken);
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker crahist
numbscopienew_e_cat / RL;
TITLE 'Crude model';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= v85aa age sexe smoker crahist
numbscopienew_e_cat / RL;
TITLE 'Crude model multivit';
ID famnr;
RUN;

PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= v87aa age sexe smoker crahist
numbscopienew_e_cat / RL;
TITLE 'Crude model vit C';
ID famnr;
RUN;

* Voor overzicht aantal cases tabel;
PROC SORT DATA=final_any;
BY CRAcase;
RUN;

PROC FREQ DATA=final_any;
TABLES suppluser;
BY CRAcase;
TITLE 'any supplement';
RUN;
```

```
PROC FREQ DATA=final_any;
TABLES v85aa;
BY CRAcase;
TITLE 'multivit';
RUN;

PROC FREQ DATA=final_any;
TABLES v87aa;
BY CRAcase;
TITLE 'vit C';
RUN;

PROC FREQ DATA=final_any;
TABLES v92aa;
BY CRAcase;
TITLE 'calcium';
RUN;

PROC FREQ DATA=final_any;
TABLES v94aa;
BY CRAcase;
TITLE 'visolie';
RUN;

PROC FREQ DATA=final_any;
TABLES v95aa;
BY CRAcase;
TITLE 'other';
RUN;

* Voor overzicht perstime in tabel;
PROC SORT DATA=final_any;
BY suppluser;
RUN;

PROC UNIVARIATE DATA=final_any;
 VAR perstime_f;
 BY suppluser;
 WHERE CRAcase=1;
 TITLE 'any supplement';
 RUN;

PROC SORT DATA=final_any;
BY v85aa;
RUN;

PROC UNIVARIATE DATA=final_any;
 VAR perstime_f;
 BY v85aa;
```

```
 WHERE CRAcase=1;
 TITLE 'multivit';
 RUN;

PROC SORT DATA=final_any;
BY v87aa;
RUN;

PROC UNIVARIATE DATA=final_any;
 VAR perstime_f;
 BY v87aa;
 WHERE CRAcase=1;
 TITLE 'vit c';
 RUN;

PROC SORT DATA=final_any;
BY v92aa;
RUN;

 PROC UNIVARIATE DATA=final_any;
 VAR perstime_f;
 BY v92aa;
 WHERE CRAcase=1;
 TITLE 'calcium';
 RUN;

PROC SORT DATA=final_any;
BY v94aa;
RUN;

PROC UNIVARIATE DATA=final_any;
 VAR perstime_f;
 BY v94aa;
 WHERE CRAcase=1;
 TITLE 'visolie';
 RUN;

PROC SORT DATA=final_any;
BY v95aa;
RUN;

PROC UNIVARIATE DATA=final_any;
 VAR perstime_f;
 BY v95aa;
 WHERE CRAcase=1;
 TITLE 'other';
 RUN;
```

```
* Testen andere definitie supplement user (final_wk) (niet getest of daar
evt. wel confounding is);
* Crude model;
* Any supplement;
PROC PHREG DATA=final_wk covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
/ RL;
TITLE 'Crude model any use';
ID famnr;
RUN;

* Multivitamin;
PROC PHREG DATA=final_wk covs(aggregate);
MODEL perstime_f*CRAcase(0)= v85aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model multivit';
ID famnr;
RUN;

* Vitamin C;
PROC PHREG DATA=final_wk covs(aggregate);
MODEL perstime_f*CRAcase(0)= v87aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model vit C';
ID famnr;
RUN;

* Calcium;
PROC PHREG DATA=final_wk covs(aggregate);
MODEL perstime_f*CRAcase(0)= v92aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model calcium';
ID famnr;
RUN;

* Visolie;
PROC PHREG DATA=final_wk covs(aggregate);
MODEL perstime_f*CRAcase(0)= v94aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model visolie';
ID famnr;
RUN;

* Other;
PROC PHREG DATA=final_wk covs(aggregate);
MODEL perstime_f*CRAcase(0)= v95aa age sexe smoker numbscopienew_e_cat /
RL;
TITLE 'Crude model other';
```

```
ID famnr;
RUN;

* Voor tabel overzicht aantal cases en HR;
PROC SORT DATA=final_wk;
BY CRAcase;
RUN;

PROC FREQ DATA=final_wk;
TABLES suppluser;
BY CRAcase;
TITLE 'any supplement';
RUN;

PROC FREQ DATA=final_wk;
TABLES v85aa;
BY CRAcase;
TITLE 'multivit';
RUN;

PROC FREQ DATA=final_wk;
TABLES v87aa;
BY CRAcase;
TITLE 'vit C';
RUN;

PROC FREQ DATA=final_wk;
TABLES v92aa;
BY CRAcase;
TITLE 'calcium';
RUN;

PROC FREQ DATA=final_wk;
TABLES v94aa;
BY CRAcase;
TITLE 'visolie';
RUN;

PROC FREQ DATA=final_wk;
TABLES v95aa;
BY CRAcase;
TITLE 'other';
RUN;

* Testen Proportional hazard assumption;
PROC PHREG DATA=final_any covs(aggregate);
MODEL perstime_f*CRAcase(0)= suppluser age sexe smoker numbscopienew_e_cat
/ RL;
TITLE 'Crude model any use';
```

```
ID famnr;
ASSESS <suppluser> <PH>;
RUN;

PROC LIFETEST METHOD=KM PLOTS=(LLS) DATA=final_any;
TIME perstime_f*CRAcase(0);
STRATA suppluser;
RUN;

* Vanaf hier niks mee gedaan;

* 1Jan1900 is missing value (niet nodig, want als er ergens 01Jan1900 stond
ingevuld (bijv. bij dtscltst) dan had die
persoon bijv. cancer, dus daarom niet ingevuld...;
DATA ES.CRTcases2;
SET ES.CRTcases1;
IF dtscltst_1=01JAN1900 THEN dtscltst_1='.';
IF dtscltst_2 =01JAN1900 THEN dtscltst_2='.';
IF dtscltst_3=01JAN1900 THEN dtscltst_3='.';
IF dtscltst_4=01JAN1900 THEN dtscltst_4='.';
IF dtscltst_5=01JAN1900 THEN dtscltst_5='.';
IF datediag_1=01JAN1900 THEN datediag_1='.';
IF datediag_2=01JAN1900 THEN datediag_2='.';
IF datediag_3=01JAN1900 THEN datediag_3='.';
IF datediag_4=01JAN1900 THEN datediag_4='.';
IF datediag_5=01JAN1900 THEN datediag_5='.';
IF date_pa_1=01JAN1900 THEN date_pa_1='.';
IF date_pa_2=01JAN1900 THEN date_pa_2='.';
IF date_pa_3=01JAN1900 THEN date_pa_3='.';
IF date_pa_4=01JAN1900 THEN date_pa_4='.';
IF date_pa_5=01JAN1900 THEN date_pa_5='.';
RUN;

* Aanmaken persontime voor adenomatous polyp, cancer and death cases.
Ga ervan uit dat SAS 'terugtelt', dus de mensen die CRAcaseFU1 én
CRAcaseFU2 zijn,
krijgen de perstime_fu_adenoom voor 1e scopie
Zie syntax hieronder, klopte niet, SAS telt niet terug. Had ELSE IF moeten
gebruiken???;
DATA total5_fu2;
SET total5_fu1_1;
IF CRACaseFU1=1 THEN perstime_fu_adenoom=dtscltst_1-START;

IF CRACaseFU2=1 THEN perstime_fu_adenoom=dtscltst_2-START;
IF CRACaseFU3=1 THEN perstime_fu_adenoom=dtscltst_3-START;
IF CRACaseFU4=1 THEN perstime_fu_adenoom=dtscltst_4-START;
IF CRACaseFU5=1 THEN perstime_fu_adenoom=dtscltst_5-START;
IF CRACaseFU6=1 THEN perstime_fu_adenoom=dtscltst_6-START;
IF CRACaseFU7=1 THEN perstime_fu_adenoom=dtscltst_7-START;
```

```
IF CRACaseFU8=1 THEN perstime_fu_adenoom=dtscltst_8-START;

IF CRCcaseFU1=1 THEN perstime_fu_cancer=datediag_1-START;
IF CRCcaseFU2=1 THEN perstime_fu_cancer=datediag_2-START;
IF CRCcaseFU3=1 THEN perstime_fu_cancer=datediag_3-START;
IF CRCcaseFU4=1 THEN perstime_fu_cancer=datediag_4-START;
IF CRCcaseFU5=1 THEN perstime_fu_cancer=datediag_5-START;
IF CRCcaseFU6=1 THEN perstime_fu_cancer=datediag_6-START;
IF CRCcaseFU7=1 THEN perstime_fu_cancer=datediag_7-START;
IF CRCcaseFU8=1 THEN perstime_fu_cancer=datediag_8-START;

IF ovl=1 THEN perstime_fu_death=ovldt-START;
RUN;

* Check bovenstaande syntax, klopt niet, o.a. A1236, A1247 en A1252 hebben
perstime tot aan latere scopie dan 1e
poliep gekregen. Daarom syntax8 (ipv 7 aangemaakt);
DATA total5_fu2_check;
SET total5_fu2;
KEEP deelnr CRACaseFU1 CRACaseFU2 CRACaseFU3 CRACaseFU4 CRACaseFU5
CRACaseFU6 CRACaseFU7 CRACaseFU8
CRCCaseFU1 CRCCaseFU2 CRCCaseFU3 CRCCaseFU4 CRCCaseFU5 CRCCaseFU6
CRCCaseFU7 CRCCaseFU8
perstime_fu_adenoom perstime_fu_cancer perstime_fu_death START dtscltst_1
dtscltst_2 dtscltst_3 dtscltst_4
dtscltst_5 dtscltst_6 dtscltst_7 dtscltst_8 datediag_1 datediag_2
datediag_3 datediag_4 datediag_5 datediag_6
datediag_7 datediag_8;
RUN;

* Uitrekenen persontime voor noncases FU. Ervan uitgaan dat SAS terugtelt,
klopt niet, SAS telt niet terug
349 observations;
DATA total5_noncases2_2;
SET total5_noncases2_1;
IF dtscltst_8 NE '.' THEN perstime_fu_noncase=dtscltst_8-START;
IF dtscltst_7 NE '.' THEN perstime_fu_noncase=dtscltst_7-START;
IF dtscltst_6 NE '.' THEN perstime_fu_noncase=dtscltst_6-START;
IF dtscltst_5 NE '.' THEN perstime_fu_noncase=dtscltst_5-START;
IF dtscltst_4 NE '.' THEN perstime_fu_noncase=dtscltst_4-START;
IF dtscltst_3 NE '.' THEN perstime_fu_noncase=dtscltst_3-START;
IF dtscltst_2 NE '.' THEN perstime_fu_noncase=dtscltst_2-START;
IF dtscltst_1 NE '.' THEN perstime_fu_noncase=dtscltst_1-START;
RUN;


* Overview supplement intake GeoLynch study (this one didn't make any
sense);
PROC PLOT DATA=compleet_perstime_cracase;
```

```
PLOT v85b * v85a;
PLOT v87b * v87a;
PLOT v88b * v88a;
PLOT v89b * v89a;
PLOT v90b * v90a;
PLOT v91b * v91a;
PLOT v92b * v92a;
PLOT v93b * v93a;
PLOT v94b * v94a;
PLOT v95b * v95a;
RUN;

* Aantal scopies, maar werkte niet op deze manier;
DATA compleet_perstime_cracase_nrsc;
SET compleet_perstime_cracase;
FUP1=dtscltst_1-START;
FUP2=dtscltst_2-START;
FUP3=dtscltst_3-START;
FUP4=dtscltst_4-START;
FUP5=dtscltst_5-START;
FUP6=dtscltst_6-START;
FUP7=dtscltst_7-START;
FUP8=dtscltst_8-START;
RUN;

DATA compleet_perstime_cracase_nrsc2;
SET compleet_perstime_cracase_nrsc;
IF 0 < FUP1 <= perstime_f THEN scofup1new = 1 ; ELSE scofup1new = 0;
IF 0 < FUP2 <= perstime_f THEN scofup2new = 1 ; ELSE scofup2new = 0;
IF 0 < FUP3 <= perstime_f THEN scofup3new = 1 ; ELSE scofup3new = 0;
IF 0 < FUP4 <= perstime_f THEN scofup4new = 1 ; ELSE scofup4new = 0;
IF 0 < FUP5 <= perstime_f THEN scofup5new = 1 ; ELSE scofup5new = 0;
IF 0 < FUP6 <= perstime_f THEN scofup6new = 1 ; ELSE scofup6new = 0;
IF 0 < FUP7 <= perstime_f THEN scofup7new = 1 ; ELSE scofup7new = 0;
IF 0 < FUP8 <= perstime_f THEN scofup8new = 1 ; ELSE scofup8new = 0;

numbscopienew_e = (scofup1new + scofup2new + scofup3new + scofup4new +
scofup5new + scofup6new
+ scofup7new + scofup8new) ;

LABEL numbscopienew= 'number of (colo/sigmo/recto) scopies during
persoonstijd';
run;

DATA compleet_perstime_cracase_nrsc3;
SET compleet_perstime_cracase_nrsc2;
totalscopienew_e=numbscopienew_f+numbscopienew_e;
RUN;
```

```
PROC PRINT DATA=compleet_perstime_cracase_nrsc3;
VAR scofup1new scofup2new scofup3new scofup4new scofup5new scofup6new
totalscopienew_e numbscopienew_f numbscopienew_e;
RUN;
```