

COMPARISON OF RISK FACTORS FOR COLON AND RECTAL CANCER

Esther K. WEI^{1,3*}, Edward GIOVANNUCCI^{2,3}, Kana WU⁴, Bernard ROSNER^{3,5}, Charles S. FUCHS^{3,6}, Walter C. WILLETT^{2,3} and Graham A. COLDITZ^{1,3}

¹Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

²Departments of Nutrition and Epidemiology, Harvard School of Public Health, Boston, MA, USA

³Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

⁴Department of Nutrition, Harvard School of Public Health, Boston, MA, USA

⁵Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

⁶Department of Adult Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

Predictors of colorectal cancer have been extensively studied with some evidence suggesting that risk factors vary by subsite. Using data from 2 prospective cohort studies, we examined established risk factors to determine whether they were differentially associated with colon and rectal cancer. Our study population included 87,733 women from the Nurses' Health Study (NHS) and 46,632 men from the Health Professionals Follow Up Study (HPFS). Exposure information was collected via biennial questionnaires (dietary variables were collected every 4 years). During the follow-up period (NHS: 1980 to May 31, 2000; HPFS: 1986 to January 31, 2000), we identified 1,139 cases of colon cancer and 339 cases of rectal cancer. We used pooled logistic regression to estimate multivariate relative risks for the 2 outcomes separately and then used polytomous logistic regression to compare these estimates. In the combined cohort, age, gender, family history of colon or rectal cancer, height, body mass index, physical activity, folate, intake of beef, pork or lamb as a main dish, intake of processed meat and alcohol were significantly associated with colon cancer risk. However, only age and sex were associated with rectal cancer. In a stepwise polytomous logistic regression procedure, family history and physical activity were associated with statistically significant different relative risks of colon and rectal cancer. Our findings support previous suggestions that family history and physical activity are not strong contributors to the etiology of rectal cancer. Future investigations of colon or rectal cancer should take into consideration risk factor differences by subsite.

© 2003 Wiley-Liss, Inc.

Key words: colon cancer; rectal cancer; cohort studies; risk factors

Many etiologic factors have been established as contributing to the risk of colorectal cancer. However, the statistical models that evaluate the endpoint of colorectal cancer assume constancy of the relative risks for colon cancer and rectal cancer, or ignore if heterogeneity exists between them. If colon cancer and rectal cancer have distinct etiologies, these assumptions may be inappropriate and established risk factors for colorectal cancer may or may not be risk factors for colon or rectal cancer if considered separately.

Recently, several investigators have studied associations between exposures and subsites in the large bowel. These studies suggest etiologic distinctions by anatomic location within the colon.^{1–6} For example, physical inactivity and body mass index have been associated with colon cancer but not with rectal cancer.⁷ Because rectal cancers make up approximately 30% of all large bowel cancers in developed countries, analyses of rectal cancer as a distinct endpoint have often been inconclusive as a result of limited statistical power. To our knowledge, no study has formally compared risk factors for colon cancer and rectal cancer. To further our understanding of etiologic similarities and differences between colon cancer and rectal cancer, we used polytomous logistic regression techniques to examine and compare risk estimates for colon cancer and rectal cancer across a wide range of dietary and lifestyle factors.

MATERIAL AND METHODS

Study cohorts

We used data from 2 prospective cohort studies, the Nurses' Health Study (NHS) and the Health Professionals Follow-Up Study (HPFS). The Nurses' Health Study began in 1976 with 121,700 female married nurses who were 30 to 55 years of age and resided in 11 states in the United States. Every 2 years since 1976, the Nurses have been sent questionnaires to ascertain their current lifestyle habits and disease history. Dietary intake was first assessed in 1980 and subsequently in 1984, 1986 and every 4 years since 1986.

The Health Professionals Follow-Up Study is a prospective cohort study begun in 1986. At baseline, it was comprised of 51,529 male dentists, optometrists, osteopaths, podiatrists, pharmacists and veterinarians between 40 and 75 years of age in 1986. The cohort is sent a questionnaire every 2 years to update information on current lifestyle habits and disease history. Dietary intake was first assessed in 1986 and every 4 years since then. We analyzed data for men and women separately and also aggregated the data from the cohorts. However, we performed the polytomous logistic regression analysis only on the combined cohort due to sample size limitations.

Study population

We excluded any participants who reported previous cancer (except for nonmelanoma skin cancer), ulcerative colitis, Crohn's disease or familial polyposis syndrome at baseline. We also excluded individuals who reported implausible caloric intakes (<800 or >4,200 kcal for men and <600 or >3,500 kcal for women) or who had a significant number of items left blank on the FFQ (10 or more blank out of the 61 question food-frequency questionnaire in 1980 for the women, or 70 or more on the 131 item food-frequency questionnaire in 1986 for the men). This left 87,733 women and 46,632 men eligible for analysis and a combined total of 134,365 subjects in the combined cohort.

Exposure assessment

All exposure information was based on data collected via the biennial questionnaires. The primary exposures of interest defined *a priori* were family history, smoking, physical activity, body mass

Grant sponsor: NIH; Grant number: CA87969

*Correspondence to: Channing Laboratory, 181 Longwood Avenue, 3rd Floor, Boston, MA 02115. Fax: +617-525-2008.
E-mail: esther.wei@channing.harvard.edu

Received 24 January 2003; Revised 19 June 2003, 9 August 2003;
Accepted 22 August 2003

DOI 10.1002/ijc.11540

index (weight in kilograms divided by square of height in meters), height, calcium, folate, alcohol, red meat and processed meat. Each factor had been evaluated previously in relation to colon cancer in these 2 cohorts.^{5,8-20} The disparity of information collected from the NHS and HPFS questionnaires on aspirin dose and duration was such that combining the data on this variable was not feasible. Therefore, we did not include it in this analysis. We also did not consider other potential exposures such as vitamin D or alternative methods of measuring anthropometry due to a desire to limit this analysis to the risk factors with the most well-established associations. We also did not evaluate post-menopausal hormone use, due to the limited number of cases in the women.

Lifestyle variables

The smoking behavior information collected in the NHS and HPFS allows for current, past or never classification or a cumulative pack-year cigarette history. One pack-year of smoking is equivalent to having smoked 1 pack or 20 cigarettes per day for 1 year. Using information on pack-years accumulated throughout the subjects' participation in the study, we calculated pack-year totals for various intervals in the person's life. We chose to use pack-years of smoking accumulated prior to age 30 based on previous studies in both of these cohorts.^{8,9} These previous studies showed this variable largely accounted for the long latency and therefore captured the association of smoking with an increased risk for colorectal cancer. Also, using this variable facilitated use of a single continuous variable for smoking in the polytomous logistic regression models.

For physical activity in the HPFS, we divided the men into quintiles based on their MET hours per week of activity at baseline (1986). To calculate the MET hours per week, we multiplied the time spent on each activity asked about (walking or hiking outdoors, jogging, running, bicycling, lap swimming, tennis, squash or racquetball and calisthenics or rowing) by the typical energy expenditure for that activity expressed in metabolic equivalents. We also added in METS for the number of flights of stairs climbed daily and the usual walking pace. The validity of the MET-hours per week has been studied and reported previously in the NHS II cohort (a cohort similar to the NHS cohort) as well as the HPFS.^{21,22} We used the medians of each of the quintiles for the linear tests of trend and in the polytomous logistic regression.

In the NHS, we were unable to derive MET-hours comparable to the men until the 1986 questionnaire. Thus, to keep the follow-up period consistent for men and women and to use all cases that arose in the follow-up period, we used the responses on the 1980 questionnaire to classify participants into 5 physical activity categories. More specifically, in 1980 the nurses reported whether they engaged in exercise intensely enough to produce sweating and then the intensity of the specific activity that they reported engaging in most frequently. Women who reported never sweating were classified as low activity (the reference group). Women were considered to have moderate physical activity if they reported sweating and engaged in a moderately intense activity (*i.e.*, a metabolic equivalent score between 1 and 4). Participants who reported sweating and engaged in vigorous activity (MET score of 4 or more) were categorized as high activity. We further divided the moderate category into those who engaged in moderate activity less than 2 times a week or 2 or more times per week. The high activity category was also further divided into those who reported 4 to 6 METS and those with more than 6 METS. Thus, we had 5 categories based on a combination of sweating or not, and MET scores for their primary physical activity. A similar derivation of physical activity in this cohort has been published previously.²³ Starting in 1986, we were able to directly calculate a MET hours per week for each NHS participant. Therefore, in order to improve the comparability between the men and women, we used the medians of quintiles of the 1986 MET hours per week for the linear tests of trend and for the polytomous logistic regression.

Anthropometric variables

Baseline height was obtained in the NHS in 1976 and in the HPFS in 1986. Current weight was also obtained at baseline and on the biennial mailed questionnaires. We used self-reported height and weight to calculate each participant's body mass index (BMI). The self-reported height and weight variables have been previously validated in these cohorts.²⁴ After adjusting for age and within-person variability, correlations of 0.97 for the HPFS participants and 0.96 in the NHS participants were found between self-reported and measured weight (although the NHS participants underreported on average 1.5 kg).

Dietary variables

A semi-quantitative food frequency questionnaire developed specifically for these populations was used for dietary assessment in these 2 cohorts. This type of questionnaire has been extensively validated since its creation.²⁵⁻³⁰ Nutrient intakes are computed by multiplying the reported frequency of consumption of specific food items by the nutrient composition of the specified portion size, using composition values from the U.S. Department of Agriculture sources supplemented with other sources. Nutrient values are adjusted for total energy intake. For the dietary variables alcohol, folate, beef as a main dish and processed meat, we used baseline values of intakes reported (1980 for women and 1986 for men). We used baseline intake for those variables that were previously found to be associated with colon or colorectal cancer in these cohorts. For calcium, we used a cumulative average of intake over the follow-up period because this was associated with colon cancer in these cohorts.⁵

Family history

Family history of colorectal cancer was ascertained in 1982, 1988, 1992, 1996 and 2000 in the women and 1986, 1990, 1992, 1996 and 2000 for the men. We asked the nurse or health professional whether first-degree relatives had ever been diagnosed with colon or rectal cancer. If the participant responded affirmatively with a family member having been diagnosed, he or she was categorized as having a positive family history. As described in a previous analysis,³¹ we updated this information throughout the follow-up period to account for newly diagnosed family members.

Screening

For both the NHS and HPFS, the participants were asked in 1990 whether they had ever had a sigmoidoscopy or endoscopy and if so, the year they had the test done. We used the response to this question to categorize participants as ever having had an endoscopy before baseline (1980 or 1986). We controlled for screening history in all of the multivariate models as individuals with a prior endoscopy had a lower subsequent risk for colorectal cancer.

Identification of cases

For both cohorts, each mailed biennial questionnaire ascertained whether a participant was diagnosed with cancer or other disease outcomes within the previous 2 years. When a participant (or next of kin for fatal outcomes) reported a diagnosis of cancer of the colon or rectum on the questionnaire, we asked for permission to obtain hospital records and pathology reports regarding this diagnosis. A physician reviewed the medical records to verify information on histological type, anatomic location and stage of the cancer. Most deaths in these cohorts were reported by family members or the postal system. The National Death Index was used to identify deaths among nonrespondents and has been shown to be highly specific and sensitive.³²

Statistical analysis

For the age-adjusted and multivariate pooled logistic models, variables were modeled as dichotomous or in categories with the same cut points for the men and women. We chose the cut points based on previous literature but also adjusted them when necessary to keep a reasonable number of cases in each category. The only

variable for which cut points were gender-specific was height. For height, we generated 4 categories separately for men and women based on the distribution in each cohort and then combined the categories. Red meat and processed meat were modeled in servings per day.

Each study participant contributed person-time from the date of return of the 1980 questionnaire in women and the 1986 questionnaire in men to the date of a colon or rectal cancer diagnosis, death from any cause to May 31, 2000 (NHS) or January 31, 2000 (HPFS), whichever came first. The lowest category was always used as the reference group. We calculated relative risks by dividing the incidence rate in each category by the rate among those in the lowest category (reference category). We used the Mantel-Haenszel summary estimator for the stratified analysis. To test for trend, we modeled the medians from each category as a continuous variable.

To adjust for more than one variable at a time, we used logistic regression pooled over each 2 year time period. Pooled logistic regression asymptotically estimates results from a Cox proportional hazards model assuming that the time period between intervals is short and the number of event in each time period is low.³³

To test for differences in associations with colon and rectal cancer, we used polytomous logistic regression on the combined cohort.³⁴ In particular, we used a custom software program described by Marshall and Chisholm.³⁵ This program provides formal tests of the differences in magnitude of the beta estimate of each risk factor for the separate components of a composite endpoint. This program allows the user to specify which variables are modeled with a common beta estimate for both outcomes and which are modeled with 2 distinct betas.

We began with a model that allowed all of the risk estimates for each exposure variable to vary between the 2 outcomes. Then we conducted a "manual" stepwise procedure, each time constraining 1 risk factor to be uniform across the 2 endpoints while allowing all the others to vary. To evaluate the appropriateness of modeling each exposure variable as the same for both outcomes we used likelihood ratio tests. Comparing each successive model with the baseline model, we determined which variable had the highest *p*-value and set that to have one common estimate for colon and rectal cancer. This became the baseline model for the next set of models. We repeated this procedure with each of the remaining variables, setting more variables to be the same for the 2 outcomes until all the remaining variables had *p*-values less than or equal to 0.05, indicating that they were likely to have different associations with the 2 outcomes.

Using polytomous regression, modeling 3 outcome categories (nondiseased, colon cancer cases and rectal cancer cases) is extremely computationally intensive. Thus, to reduce the complexity of the computations and to simplify the interpretation of the estimates, we simplified the expressions of each variable for the polytomous logistic regression. We modeled age as a continuous variable and family history and endoscopy as dichotomous variables. For BMI, physical activity, smoking, beef as a main dish, processed meat, height, calcium, folate, alcohol and fiber, we used the medians of the categories (or quintiles) as a continuous variable. Thus, the estimates of risk from the polytomous logistic models represent risk on a different scale than the pooled logistic models and cannot be compared directly. However, when variables are defined identically, results from the polytomous logistic regression and a traditional logistic model are comparable.³⁶ For variables that had missing values (alcohol, physical activity and endoscopy), we assigned an indicator variable for "missing" and another variable for the value if not missing. In the polytomous logistic regression stepwise procedure, we moved both variables in and out of the model together.

RESULTS

Table I shows the distribution of the various exposure variables across the outcome categories. For both the men and women, we show the values at baseline for each cohort (1980 for women and 1986 for men). The combined cohort had 1,139 colon cancer cases, and 339 rectal cancer cases over 2,302,712 person years. (For the distribution of these cases by cohort, see Table I). The men were more physically active and taller, had higher intake of alcohol, had slightly higher intakes of folate, calcium and processed meats and smoked more pack-years before age 30 when compared to the women. Rectal cancer cases tended to have slightly higher folate and slightly lower calcium intake, to be more physically active and to have a lower frequency of positive family history of colorectal cancer when compared to colon cancer cases.

For each variable of interest, and for each cohort (NHS, HPFS, combined cohort), we calculated the number of cases and the multivariate relative risks of colon and rectal cancer (Table II). We also report *p*-values based on tests for trend. We found age, sex, family history, height, BMI, physical activity, folate, beef, pork or lamb as a main dish, processed meat and alcohol to be related to colon cancer, whereas only age and gender showed statistically significant associations with rectal cancer.

For BMI, there was a suggestion of a different effect between men and women. Although there was an increased risk for colon cancer for both men and women, there was also a statistically significant increased risk in the highest category of BMI among the women for rectal cancer (MVR = 1.56, 95% confidence interval: 1.01 to 2.42; *p* for trend = 0.04).

Based on the manual stepwise polytomous logistic regression procedure, we found that family history and physical activity had significantly different risk estimates for colon versus rectal cancer. Height had the lowest nonstatistically significant *p*-value for a difference between colon cancer and rectal cancer (*p* = 0.07). Table III shows results from the polytomous logistic regression including the *p*-values for the difference between the estimates for colon cancer versus rectal cancer. Log likelihood methods indicated that all other variables appear to be modeled appropriately when risk estimates were set to be the same for colon and rectal cancer.

DISCUSSION

In a combined cohort made up 134,365 women and men, we found that relative risks for family history of colon or rectal cancer and physical activity were significantly different between colon cancer and rectal cancer.

Our finding, that physical activity is more strongly associated with colon cancer rather than rectal cancer, is consistent with the previous literature. One possible biological mechanism to explain this difference is that the colon is more susceptible to the effects of insulin. With increased physical activity, insulin sensitivity improves.³⁷ If insulin and insulin-like growth factors are responsible for promoting adenoma development or growth, it is possible that the receptors responsible for this effect are less common in the rectum compared to the colon. Furthermore, we found some disparity between the effect of BMI in men and women. These results are consistent with the hypothesis that the mechanism by which BMI influences risk of colon and rectal cancer are more complex and less consistent in women than in men, possibly influenced by the effects of estrogen.³⁸

Similarly, our findings of an increased risk for colon cancer with increased adult height are consistent with previous literature.¹³ Previously, it has been proposed that adult height may be a proxy for energy balance during childhood, which would also be related to insulin resistance and levels of growth factors.¹¹ As proposed above, one explanation for the differential association we found for height and physical activity between the colon and rectum is the possibility that the colon has a different pattern of receptors for these growth factors and/or a different sensitivity to insulin effects compared to the rectum. The potential for a gradient of hormone

TABLE 1—AGE-STANDARDIZED DISTRIBUTION OF RISK FACTORS FOR COLON AND RECTAL CANCER AT BASELINE FOR WOMEN (1980) AND MEN (1986)

		Noncases 86,857 46,030	Colon cancer 672 467	Rectal cancer 204 135
Mean				
BMI (kg/m ²)				
	Women	24.5	24.7	25.4
	Men	25.5	25.8	24.9
Physical Activity (MET hours per week)				
	Women	13.4	12.1	13.5
	Men	20.9	17.8	20.2
Height (inches)				
	Women	64.5	64.9	64.2
	Men	70.1	70.6	70.3
Alcohol (gm/day)				
	Women	6.4	6.6	8.2
	Men	11.3	14.1	13.8
Total Folate (mcg/day)				
	Women	365	356	365
	Men	480	466	471
Total Calcium (mg/day)				
	Women	733	712	708
	Men	898	889	872
Beef, pork or lamb as a main dish (serv/day)				
	Women	0.37	0.39	0.35
	Men	0.31	0.30	0.26
Processed meat (serv/day)				
	Women	0.17	0.20	0.18
	Men	0.26	0.23	0.25
Pack-years of smoking before age 30				
	Women	3.0	3.2	3.4
	Men	5.4	6.1	6.0
Percent of group				
Female		73.9	58.7	59.9
Positive family history				
	Women	7.8	14.5	14.4
	Men	8.4	13.8	12.0

receptors between the colon and rectum is supported by the fact that the colon and rectal mucosa arise from different embryonic tissue, the colon arising from the midgut and the rectum arising from the hindgut.⁶

More generally, the colon and rectum serve different functions and are exposed to fecal matter for different durations. The rectum is exposed to the fecal in a more concentrated and direct way compared to the colon. Also, as undigested matter travels through the colon, it is coated with alkaline mucus. The different levels of pH in proximal and distal locations within the colon may influence susceptibility to environmental factors.⁶

In their histologic study of large bowel carcinogenesis, Konishi *et al.*³⁹ suggest that cancers that arise in the distal colon and rectum begin as adenomatous polyps whereas a *de novo* pathway is more important in lesions that arise in the proximal colon. Also, the prevalence of K-ras mutations and mutation patterns in the *p53* gene in rectal cancers differ from those seen in colon cancers.⁴⁰ In a recent review, Iocapetta highlights the clinical and molecular differences between tumors of the proximal and distal colon. One compelling difference is that the familial forms of colorectal cancer, familial polyposis syndrome (FAP) and hereditary nonpolyposis coli (HNPCC), arise first in different sections of the colon (FAP: rectum and distal colon, HNPCC: proximal colon).⁶ These findings are compatible with different etiologic factors for colon cancer and rectal cancers.

Previously, risk estimates for rectal cancer have been reported in the NHS and the HPFS cohorts for several variables. These secondary analyses have suggested that there are etiologic differences between the 2 outcomes. For example, in an analysis of family history and colorectal cancer in these 2 cohorts, Fuchs *et al.*³¹

reported relative risk (RR) of 1.99 (95% confidence interval (CI): 1.50–2.63) for colon cancer and a relative risk of 0.86 (95% CI: 0.43–1.70) for rectal cancer among those who reported a family history of colon or rectal cancer. They report a *p*-value for the difference between colon and rectal cancer that is based on a simple 1 degree of freedom chi-square test. Our methods offer a more rigorous comparison of these risk estimates for the 2 outcomes and provide flexibility in allowing some risk estimates to be the same and other to be different based on likelihood ratio methods.³⁵ Our results confirm the hypothesis that a family history of colon or rectal cancer appears to affect risk for colon cancer more strongly than risk for rectal cancer; however, because the questionnaire only queried a history of colon *or* rectal cancer, we cannot rule out the possibility that colon cancer cases arose in those with a family history of colon cancer and rectal cancer arose in those with a family history of rectal cancer.

For several variables such as smoking, calcium and beef as a main dish, our results appear slightly weaker than what was previously reported in these cohorts. Generally, these differences can be explained by comparing model specification and follow-up time from each of those results. Previous analyses within these cohorts often used covariates in their models that were at the time known or suspected to be confounders. With time, our understanding of possible confounding factors has improved and therefore we are now able to control for confounding more precisely. Thus, in comparison to prior analyses we have increased the number of covariates in our models and hopefully better captured the underlying associations. Furthermore, with longer follow-up time and therefore more cases, we have improved power to report accurate and precise risk estimates, particularly with respect to rectal can-

TABLE II—RELATIVE RISKS FOR COLON AND RECTAL CANCER IN 87,733 WOMEN IN THE NURSES' HEALTH STUDY AND 46,632 MEN IN THE HEALTH PROFESSIONALS FOLLOW-UP STUDY

IN THE HEALTH PROFESSIONALS FOLLOW-UP STUDY						
Sex	Categories					
	Male		Female			
Cases						
Colon		467			672	
Rectum		135			204	
MVRR (95% CI) ^{1,2}						
Colon		1.00			0.63 (0.54, 0.73)	
Rectum		1.00			0.71 (0.54, 0.95)	
Family history of colorectal cancer		Categories				
		No	Yes			
NHS						
Cases						
Colon		550			122	
Rectum		176			28	
MVRR (95%CI)						
Colon		1.00			1.86 (1.52, 2.26)	
Rectum		1.00			1.33 (0.89, 1.99)	
HPFS						
Cases						
Colon		378			89	
Rectum		119			16	
MVRR (95%CI)						
Colon		1.00			1.69 (1.34, 2.14)	
Rectum		1.00			1.00 (0.59, 1.69)	
COMBINED COHORT						
Cases						
Colon		928			211	
Rectum		295			44	
MVRR (95%CI)						
Colon		1.00			1.81 (1.55, 2.10)	
Rectum		1.00			1.20 (0.87, 1.65)	
Height (inches)		Categories				<i>p</i> ³
		<62	63–64	65–66	67–81	
NHS						
Cases						
Colon		131	193	188	160	
Rectum		42	71	53	38	
MVRR (95%CI)						
Colon		1.00	1.18 (0.94, 1.48)	1.22 (0.97, 1.52)	1.48 (1.18, 1.88)	0.001
Rectum		1.00	1.31 (0.89, 1.92)	1.05 (0.70, 1.58)	1.08 (0.70, 1.69)	0.96
		<68	69–70	71–72	73–81	
HPFS						
Cases						
Colon		123	131	128	85	
Rectum		38	35	35	26	
MVRR (95%CI)						
Colon		1.00	1.11 (0.87, 1.43)	1.28 (1.00, 1.65)	1.50 (1.13, 2.00)	0.004
Rectum		1.00	0.92 (0.58, 1.46)	1.07 (0.68, 1.71)	1.42 (0.85, 2.35)	0.24
		1	2	3	4	
COMBINED COHORT						
Cases						
Colon		214	286	270	214	
Rectum		73	96	76	53	
MVRR (95%CI)						
Colon		1.00	1.16 (0.98, 1.37)	1.24 (1.04, 1.46)	1.50 (1.25, 1.79)	<.0001
Rectum		1.00	1.14 (0.85, 1.52)	1.04 (0.76, 1.41)	1.18 (0.84, 1.64)	0.43

cer. Because we combined the results from the men and women and did not evaluate finer definitions of subsite (*e.g.*, proximal colon vs. distal colon), we may have, for example, also underestimated associations that exist most strongly in the proximal colon.

One example of differences between our results and previous analyses in these cohorts is for pack-years before age 30. Giovannucci *et al.*⁸ has previously reported results for smoking before age 30 in these cohorts by subsite. In the HPFS, they reported a

significant association with colon cancer for ≥ 16 pack-years before age 30 vs. 0 pack-years before age 30 (RR = 1.96, 95% CI: 1.16–3.29) and found suggestive but not statistically significant results for rectal cancer (RR = 1.62, 95% CI: 0.60–4.37). These results however were based on total colorectal cases up to 1992, including only 44 rectal cancer cases. In the NHS, they reported elevated risks for both colon and rectal cancer among women 55 and older. However, the RR and CI were slightly stronger for

TABLE II—RELATIVE RISKS FOR COLON AND RECTAL CANCER IN 87,733 WOMEN IN THE NURSES' HEALTH STUDY AND 46,632 MEN IN THE HEALTH PROFESSIONALS FOLLOW-UP STUDY (CONTINUED)

Body mass index (kg/m ²)	Categories				<i>p</i>
	<23	23–24.9	25–29.9	>30	
NHS					
Cases					
Colon	210	141	207	113	
Rectum	56	46	68	34	
MVRR (95%CI)					
Colon	1.00	1.10 (0.88, 1.36)	1.11 (0.91, 1.35)	1.28 (1.10, 1.62)	0.05
Rectum	1.00	1.37 (0.92, 2.02)	1.40 (0.98, 2.01)	1.56 (1.01, 2.42)	0.04
HPFS					
Cases					
Colon	57	119	225	51	
Rectum	24	42	55	11	
MVRR (95%CI)					
Colon	1.00	1.33 (0.97, 1.83)	1.54 (1.15, 2.07)	1.85 (1.26, 2.72)	0.001
Rectum	1.00	1.16 (0.70, 1.94)	0.93 (0.57, 1.53)	1.03 (0.49, 2.14)	0.70
COMBINED COHORT					
Cases					
Colon	267	260	432	164	
Rectum	80	88	123	45	
MVRR (95%CI)					
Colon	1.00	1.14 (0.96, 1.36)	1.21 (1.04, 1.42)	1.39 (1.14, 1.69)	0.001
Rectum	1.00	1.33 (0.98, 1.81)	1.21 (0.90, 1.62)	1.40 (0.96, 2.03)	0.11

Physical activity (quintiles of MET hrs or equivalent) ⁴	Categories					<i>p</i>
	1	2	3	4	5	
NHS						
Cases						
Colon	171	189	120	113	49	
Rectum	46	48	41	37	24	
MVRR (95%CI)						
Colon	1.00	1.08 (0.88, 1.33)	1.00 (0.79, 1.26)	0.92 (0.72, 1.17)	0.73 (0.53, 1.01)	0.02
Rectum	1.00	0.99 (0.66, 1.49)	1.24 (0.81, 1.90)	1.04 (0.67, 1.62)	1.28 (0.77, 2.12)	0.35
HPFS						
Cases						
Colon	113	95	97	91	70	
Rectum	23	26	25	25	36	
MVRR (95%CI)						
Colon	1.00	0.83 (0.63, 1.09)	0.92 (0.70, 1.20)	0.77 (0.57, 1.03)	0.71 (0.52, 0.96)	0.04
Rectum	1.00	0.55 (0.30, 0.98)	0.79 (0.47, 1.34)	1.01 (0.61, 1.68)	0.95 (0.56, 1.60)	0.37
COMBINED COHORT						
Cases						
Colon	284	286	224	192	120	
Rectum	77	66	68	68	52	
MVRR (95%CI)						
Colon	1.00	0.99 (0.84, 1.17)	0.99 (0.83, 1.18)	0.87 (0.72, 1.05)	0.75 (0.61, 0.94)	0.001
Rectum	1.00	0.82 (0.59, 1.14)	1.06 (0.76, 1.48)	1.07 (0.77, 1.49)	1.14 (0.79, 1.65)	0.22

Total folate (mcg/day)	Categories				<i>p</i>
	≤200	201–300	301–400	>400	
NHS					
Cases					
Colon	168	232	97	175	
Rectum	35	79	30	60	
MVRR (95%CI)					
Colon	1.00	0.95 (0.78, 1.17)	0.86 (0.66, 1.12)	0.82 (0.66, 1.03)	0.04
Rectum	1.00	1.54 (1.03, 2.31)	1.28 (0.78, 2.12)	1.32 (0.86, 2.05)	0.94
HPFS					
Cases					
Colon	20	99	132	216	
Rectum	5	30	37	63	
MVRR (95%CI)					
Colon	1.00	0.73 (0.45, 1.19)	0.74 (0.46, 1.20)	0.72 (0.45, 1.16)	0.57
Rectum	1.00	0.80 (0.31, 2.07)	0.70 (0.27, 1.82)	0.67 (0.26, 1.72)	0.43
COMBINED COHORT					
Cases					
Colon	188	331	229	391	
Rectum	40	109	67	123	
MVRR (95%CI)					
Colon	1.00	0.91 (0.75, 1.09)	0.85 (0.69, 1.04)	0.82 (0.68, 0.99)	0.06
Rectum	1.00	1.39 (0.96, 2.01)	1.15 (0.76, 1.74)	1.18 (0.80, 1.74)	0.83

TABLE II – RELATIVE RISKS FOR COLON AND RECTAL CANCER IN 87,733 WOMEN IN THE NURSES' HEALTH STUDY AND 46,632 MEN IN THE HEALTH PROFESSIONALS FOLLOW-UP STUDY (CONTINUED)

Total calcium (mg/day)	Categories					<i>p</i>
	<600	601–700	701–900	901–1,100	>1,100	
NHS						
Cases						
Colon	147	93	172	126	134	
Rectum	48	23	56	32	45	
MVRR (95%CI)						
Colon	1.00	1.09 (0.84, 1.42)	1.06 (0.85, 1.34)	1.04 (0.81, 1.33)	0.89 (0.69, 1.15)	0.25
Rectum	1.00	0.75 (0.46, 1.24)	0.93 (0.63, 1.39)	0.67 (0.42, 1.07)	0.78 (0.50, 1.21)	0.25
HPFS						
Cases						
Colon	102	68	113	66	118	
Rectum	22	21	33	23	36	
MVRR (95%CI)						
Colon	1.00	0.87 (0.64, 1.19)	0.78 (0.59, 1.03)	0.75 (0.54, 1.03)	0.90 (0.68, 1.20)	0.72
Rectum	1.00	1.23 (0.67, 2.25)	1.10 (0.58, 1.76)	1.21 (0.66, 2.22)	1.30 (0.74, 2.27)	0.34
COMBINED COHORT						
Cases						
Colon	249	161	285	192	252	
Rectum	70	44	89	55	81	
MVRR (95%CI)						
Colon	1.00	0.99 (0.81, 1.22)	0.94 (0.79, 1.12)	0.93 (0.76, 1.13)	0.88 (0.73, 1.07)	0.17
Rectum	1.00	0.89 (0.60, 1.30)	0.94 (0.68, 1.30)	0.83 (0.58, 1.21)	0.92 (0.65, 1.30)	0.66
Beef, pork or lamb as a main dish (servings/day)	Categories					<i>p</i>
	0	<3 per month	1 per week	2 to 4 per week	5 or more per week	
NHS						
Cases						
Colon	13	70	194	270	123	
Rectum	4	24	58	93	24	
MVRR (95%CI)						
Colon	1.00	1.22 (0.67, 2.22)	1.16 (0.66, 2.05)	1.20 (0.68, 2.11)	1.31 (0.73, 2.36)	0.44
Rectum	1.00	1.37 (0.47, 3.97)	1.14 (0.41, 3.19)	1.37 (0.50, 3.80)	0.92 (0.31, 2.71)	0.52
HPFS						
Cases						
Colon	30	100	139	166	32	
Rectum	10	27	42	49	7	
MVRR (95%CI)						
Colon	1.00	1.53 (1.01, 2.33)	1.35 (0.89, 2.03)	1.44 (0.96, 2.18)	1.35 (0.80, 2.27)	0.70
Rectum	1.00	1.27 (0.61, 2.66)	1.15 (0.56, 2.38)	1.27 (0.62, 2.63)	0.90 (0.34, 2.45)	0.86
COMBINED COHORT						
Cases						
Colon	43	170	333	436	155	
Rectum	14	51	100	142	31	
MVRR (95%CI)						
Colon	1.00	1.43 (1.02, 2.02)	1.29 (0.92, 1.79)	1.37 (0.98, 1.90)	1.43 (1.00, 2.05)	0.25
Rectum	1.00	1.28 (0.70, 2.34)	1.13 (0.64, 2.02)	1.30 (0.73, 2.32)	0.90 (0.47, 1.75)	0.55
Processed meat (servings/day)	Categories					<i>p</i>
	0	<3 per month	1 per week	2 to 4 per week	5 or more per week	
NHS						
Cases						
Colon	220	169	125	110	44	
Rectum	67	61	34	33	7	
MVRR (95%CI)						
Colon	1.00	0.96 (0.78, 1.18)	1.10 (0.81, 1.26)	1.21 (0.95, 1.53)	1.32 (0.95, 1.83)	0.02
Rectum	1.00	1.13 (0.80, 1.61)	0.89 (0.58, 1.36)	1.18 (0.77, 1.81)	0.73 (0.33, 1.59)	0.77
HPFS						
Cases						
Colon	122	144	89	75	37	
Rectum	35	34	37	21	8	
MVRR (95%CI)						
Colon	1.00	1.01 (0.79, 1.30)	0.95 (0.71, 1.27)	0.96 (0.71, 1.30)	1.27 (0.87, 1.85)	0.37
Rectum	1.00	0.88 (0.54, 1.44)	1.53 (0.94, 2.52)	1.05 (0.59, 1.86)	1.06 (0.48, 2.33)	0.87
COMBINED COHORT						
Cases						
Colon	342	313	214	185	81	
Rectum	102	95	71	54	15	
MVRR (95%CI)						
Colon	1.00	1.02 (0.87, 1.19)	1.02 (0.85, 1.21)	1.13 (0.94, 1.37)	1.33 (1.04, 1.70)	0.008
Rectum	1.00	1.05 (0.79, 1.40)	1.14 (0.83, 1.56)	1.15 (0.82, 1.63)	0.90 (0.52, 1.57)	0.93

TABLE II—RELATIVE RISKS FOR COLON AND RECTAL CANCER IN 87,733 WOMEN IN THE NURSES' HEALTH STUDY AND 46,632 MEN IN THE HEALTH PROFESSIONALS FOLLOW-UP STUDY (CONTINUED)

Alcohol (grams/day)	Categories					<i>p</i>
	0	<10	10 to 19	≥20	Past	
NHS						
Cases						
Colon	200	281	106	69	16	
Rectum	56	91	28	24	5	
MVRR (95%CI)						
Colon	1.00	0.97 (0.81, 1.17)	0.99 (0.78, 1.27)	1.14 (0.86, 1.52)	0.64 (0.38, 1.07)	0.27
Rectum	1.00	1.12 (0.80, 1.58)	0.96 (0.60, 1.54)	1.48 (0.90, 2.44)	0.72 (0.29, 1.81)	0.23
HPFS						
Cases						
Colon	37	149	98	111	72	
Rectum	11	43	35	28	18	
MVRR (95%CI)						
Colon	1.00	1.08 (0.75, 1.55)	1.27 (0.86, 1.86)	1.55 (1.05, 2.27)	1.32 (0.89, 1.98)	0.003
Rectum	1.00	0.93 (0.47, 1.82)	1.31 (0.66, 2.62)	1.11 (0.54, 2.29)	1.06 (0.50, 2.27)	0.60
COMBINED COHORT						
Cases						
Colon	237	430	204	180	88	
Rectum	67	134	63	52	23	
MVRR (95%CI)						
Colon	1.00	0.97 (0.82, 1.14)	1.04 (0.85, 1.26)	1.27 (1.03, 1.56)	1.02 (0.79, 1.32)	0.003
Rectum	1.00	1.04 (0.77, 1.40)	1.07 (0.75, 1.55)	1.26 (0.85, 1.87)	0.93 (0.56, 1.54)	0.20
Pack-years smoked before age 30	Categories					<i>p</i>
	0		1 to 10		>10	
NHS						
Cases						
Colon		302		256		91
Rectum		99		77		27
MVRR (95%CI)						
Colon		1.00		0.97 (0.82, 1.15)		1.30 (1.03, 1.66)
Rectum		1.00		0.91 (0.67, 1.23)		1.18 (0.76, 1.82)
HPFS						
Cases						
Colon		197		136		117
Rectum		44		42		38
MVRR (95%CI)						
Colon		1.00		1.00 (0.80, 1.25)		1.04 (0.82, 1.32)
Rectum		1.00		1.42 (0.92, 2.19)		1.67 (1.07, 2.62)
COMBINED COHORT						
Cases						
Colon		499		392		208
Rectum		143		119		65
MVRR (95%CI)						
Colon		1.00		0.99 (0.86, 1.13)		1.13 (0.95, 1.34)
Rectum		1.00		1.07 (0.83, 1.37)		1.30 (0.96, 1.77)

¹MVRR: Multivariate relative risk. Multivariate model adjusted for age, family history, bmi, physical activity, beef, pork or lamb as a main dish, processed meat, alcohol, calcium, folate, height, pack-years smoking before age 30, history of endoscopy and gender in combined cohort.—²95% CI: 95% confidence interval.—³*p* based on test for trend.—⁴Physical activity categories in NHS calculated based on sweating or not and METS of primary activity reported in 1980. Medians of MET hr reported in 1986 used for test for trend. In HPFS, categories are quintiles of MET hr/week reported in 1986. Medians of quintiles used for test of trend.

rectal cancer (greater than 10 pack-years vs. zero pack-years: RR colon = 1.59, 95% CI: 1.14–2.23; RR rectal = 2.36, 95% CI: 1.22–4.55).⁹ Although our results showed weaker associations for colon cancer and rectal cancer than what was found previously, our relative risks are within the confidence intervals they reported. Furthermore, their multivariate models were specified with different covariates than in our study. Giovannucci *et al.*⁸ controlled for age, family history, BMI, and intake of saturated fat, dietary fiber, folate and alcohol, whereas our multivariate models also include height, beef, pork or lamb as a main dish, processed meat, calcium and history of endoscopy. Our results that show a slightly stronger association between cigarette smoking and rectal cancer than for colon cancer are also consistent with a recent prospective cohort study by Terry *et al.*⁴¹ In their cohort of 89,835 women, they found statistically significant associations between 40 or more years since smoking commenced vs. never smokers and rectal cancer

(hazard ratio = 2.27, 95% CI: 1.06–4.87) but weaker results and nonstatistically significant associations for colon cancer (hazard ratio = 1.12, 95% CI: 0.62–2.04). Further investigations of smoking and rectal cancer should be undertaken to clarify and confirm this association.

Another variable for which our results differ slightly from the previously published findings is calcium. In their analysis, Wu *et al.*⁵ found the strongest association of calcium with disease arising in the distal colon (multivariate pooled RR for >1,250 mg/day versus = 500 mg/day = 0.65, 95% CI = 0.43–0.98). However, their results were stronger in men than women (men: multivariate RR = 0.58, 95% CI = 0.32–1.05; women: multivariate RR = 0.73, 95% CI = 0.41–1.27). Furthermore, there appeared to be a threshold effect with intakes above 700 mg/day. Since we are combining the proximal and distal colon and, in the PLR model,

TABLE III—MULTIVARIATE RELATIVE RISKS FOR COLON AND RECTAL CANCER FROM THE POLYTOMOUS LOGISTIC REGRESSION MODEL WITH CORRESPONDING *p*-VALUE FOR THE DIFFERENCE BETWEEN COLON AND RECTAL¹

	ENDPOINT		<i>p</i> for difference
	Colon	Rectum	
Positive family history	1.94 (1.80, 2.10)	1.27 (1.08, 1.50)	0.02
Physical activity (per 5 MET increase)	0.90 (0.88, 0.93)	1.02 (0.98, 1.06)	0.02
Height (per 5 inches)		1.08 (1.05, 1.12)	0.07
Servings of processed meat (one additional serving per week)		1.20 (1.06, 1.37)	0.17
Female gender		0.59 (0.54, 0.63)	0.27
Servings of beef, pork or lamb (one additional serving per week)		1.12 (0.99, 1.27)	0.32
Total folate (per 100 mcg)		0.83 (0.70, 0.97)	0.45
Total calcium (per 100 mg)		0.76 (0.47, 1.25)	0.47
Pack-years of smoking before age 30 (per 5 pack years)		1.07 (1.05, 1.10)	0.61
BMI (per 5 unit increase in BMI)		1.06 (1.03, 1.10)	0.75
Alcohol (per 5 grams)		1.05 (1.03, 1.06)	0.76
Age (per 5 years)		1.21 (1.19, 1.23)	0.92

¹Model includes all of the above risk factors as well as history of endoscopy. Relative risks presented once were fixed to be equal for the 2 outcomes based on the stepwise approach to this analysis. See text for more detail.

looking at calcium as a continuous variable, our effect estimates appear weaker. Since previously there was a suggestion of a threshold effect, the interpretation of the relative risk in the polytomous model for calcium in particular, but also to some degree the other variables, must be interpreted with caution, since for the PLR analysis we specified the exposures as continuous variables.

A visual comparison of relative risks can suggest a difference in their contribution to disease in each subsite. However, visually comparing the relative risks from separate logistic models, particularly from different cohorts can be problematic due to the dependence of standardized regression coefficients on study-specific variability in risk factors. Also, a comparison of statistical significance between outcomes is limited by the dependence of levels of statistical significance on numbers of events for each component of the outcome. One major strength of our study is that the Marshall and Chisholm method we used for the polytomous logistic regression models allows for some exposure variables to vary by outcome while others remain the same.

One limitation of our study is that with 339 rectal cancer cases, we may not have enough power to detect small differences in risk estimates between colon and rectal cancer. Therefore, we cannot rule out etiologic differences between the colon and rectum that were suggested by nonsignificant *p*-values. Given our sample size, for example, it would be premature to conclude a lack of differential risk for colon and rectal cancer with increased height (*p*=0.07).

Another limitation of our study is that we considered the colon in its entirety when in fact, etiologic distinctions between the proximal and distal colon may exist. Possibly more significant differences may lie between the proximal (or distal) colon and the rectum, but these differences are not detectable when combining estimates between the proximal and distal colon. In fact, hetero-

geneity in tumor formation and growth are likely more precisely defined by molecular characteristic rather than anatomic location. However, differentiating risk factors using anatomic definitions may undercover patterns for future analyses using molecular markers.

Some degree of measurement error in our dietary and lifestyle variables is inevitable but any random misclassification of the exposure variables would only attenuate the associations we observed and would not differentially influence one outcome and not the other. However, because in this analysis we are focused on differences between the colon and rectum, any measurement error would reduce our power to compare estimates. The questions used to obtain dietary and lifestyle data from these 2 cohorts are very similar for all variables with the exception of physical activity (and aspirin, which we did not include for this reason). The slight differences in the way we used the physical activity information may have caused some error associated with combining the cohort information for this variable. However, because we used the information from 1986 for both cohorts (directly comparable derivations of MET-hours) for the polytomous logistic regression analysis, these results should still be valid.

In summary, our findings support the hypothesis that some risk factors, including family history, physical activity and possibly height, differ in their association with colon and rectal cancer. We cannot exclude the possibility that the remaining variables we examined actually have different associations with colon and rectal cancer that were not appreciated due to lack of statistical power. Because risk factors for colorectal cancer do not appear to contribute equally to colon and rectal cancers, future investigations into risk for colorectal cancer should ideally be done differentially by subsite.

REFERENCES

- Chyou P, Nomura A, Stemmermann G. A prospective study of colon and rectal cancer among Hawaii Japanese Men. *Ann Epidemiol* 1996; 6:276–82.
- Bonithon-Kopp C, Benhamiche A. Are there several colorectal cancers? Epidemiological data. *Eur J Cancer Prevention* 1999;8: S3–12.
- Yoo K, Tajima K, Inoue M, Takezaki T, Hirose K, Hamajima N, Park SK, Kang DH, Kato T, Hirai T. Reproductive factors related to the risk of colorectal cancer by subsite: a case-control analysis. *Br J Cancer* 1999;79:1901–6.
- Young TB, Wolf DA. Case-control study of proximal and distal colon cancer and diet in Wisconsin. *Int J Cancer* 1988;42:167–75.
- Wu K, Willett WC, Fuchs CS, Colditz GA, Giovannucci EL. Calcium Intake and risk of colon cancer in women and men. *J Nat Cancer Inst* 2002;94:437–46.
- Iacopetta B. Are there two sides to colorectal cancer? *Int J Cancer* 2002;101:403–8.
- Colditz GA, Cannuscio CC, Frazier AL. Physical activity and reduced risk of colon cancer: implications for prevention. *Cancer Causes Control* 1997;8:649–67.
- Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Kearney J, Willett WC. A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. men. *J Nat Cancer Inst* 1994;86:183–91.
- Giovannucci E, Colditz GA, Stampfer MJ, Colditz GA, Ascherio A, Kearney J, Willett WC. A prospective study of cigarette smoking and risk of colorectal adenoma and colorectal cancer in U.S. women. *J Nat Cancer Inst* 1994;86:192–9.
- Giovannucci E, Rimm EB, Ascherio A, Stampfer MJ, Colditz GA, Willett WC. Alcohol, low-methionine-low-folate diets, and risk of colon cancer in men. *J Nat Cancer Inst* 1995;87:265–73.
- Giovannucci E, Ascherio A, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk for colon cancer and adenoma in men. *Ann Internal Med* 1995;122:327–34.

12. Giovannucci E, Egan KM, Hunter DJ, Stampfer MJ, Colditz GA, Willett WC, Speizer FE. Aspirin and the risk of colorectal cancer in women. *New England Journal of Medicine* 1995;333:609-14.
13. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC. Intake of fat, meat, and fiber in relation to risk of colon cancer in men. *Cancer Res* 1994;54:2390-7.
14. Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC. Aspirin use and the risk for colorectal cancer and adenoma in male health professionals [see comments]. *Ann Internal Med* 1994;121:241-6.
15. Giovannucci E, Stampfer MJ, Colditz GA, Rimm EB, Trichopoulos E, Rosner BA, Speizer FE, Willett WC. Folate, methionine, and alcohol intake and risk of colorectal adenoma. *J Nat Cancer Inst* 1993;85:875-84.
16. Giovannucci E, Colditz GA, Stampfer MJ, Willett WC. Physical activity, obesity, and risk of colorectal adenoma in women (United States). *Cancer Causes Control* 1996;7:253-63.
17. Giovannucci E, Stampfer MJ, Colditz GA, Hunter DJ, Fuchs C, Rosner BA, Speizer FE, Willett WC. Multivitamin use, folate, and colon cancer in women in the Nurses' Health Study. *Ann Internal Med* 1998;129:517-24.
18. Martinez M, Giovannucci E, Colditz GA, Stampfer MJ, Hunter DJ, Speizer FE, Wing A, Willett WC. Calcium, vitamin D, and the occurrence of colorectal cancer among women. *J Nat Cancer Inst* 1996;88:1375-82.
19. Willett WC, Stampfer MJ, Colditz GA, Rosner BA, Speizer FE. Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. *N Engl J Med* 1990;323:1664-72.
20. Martinez ME, Giovannucci E, Spiegelman D, Hunter DJ, Willett WC, Colditz GA. Leisure-time physical activity, body size, and colon cancer in women. *J Nat Cancer Inst* 1997;89:948-55.
21. Wolf AM, Hunter DJ, Colditz GA, Manson JE, Stampfer MJ, Corsano KA, Rosner B, Kriska A, Willett WC. Reproducibility and validity of a self-administered physical activity questionnaire. *Int J Epidemiol* 1994;23:991-9.
22. Chasen-Taber S, Rimm EB, Stampfer MJ, Spiegelman D, Colditz GA, Giovannucci E, Ascherio A, Willett WC. Reproducibility and validity of a self-administered physical activity questionnaire for male health professionals. *Epidemiology* 1996;7:81-6.
23. Bertone ER, Willett WC, Rosner BA, Hunter DJ, Fuchs CS, Speizer FE, Colditz GA, Hankinson SE. Prospective study of recreational physical activity and ovarian cancer. *J Nat Cancer Inst* 2001;93:942-8.
24. Rimm EB, Stampfer MJ, Colditz GA, Chute CG, Litin LB, Willett W. Validity of self-reported waist and hip circumferences in men and women. *Epidemiology* 1990;1:466-473.
25. Feskansich D, Rimm E, Giovannucci E. Reproducibility and validity of a food intake measurement from semi-quantitative food frequency questionnaire. *J Am Dietetic Assoc* 1993;93:790-6.
26. Mullen B, Krantzler N, Grivetti L, Schultz H, Meiselman H. Validity of a food frequency questionnaire for the determination of individual food intake. *Am J Clin Nutr* 1984;39:136.
27. Rimm EB, Giovannucci E, Stampfer MJ, Colditz GA, Litin L, Willett WC. Reproducibility and validity of an expanded self-administered semi-quantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;135:1114-26.
28. Sampson L. Food frequency questionnaires as a research instrument. *Clin Nutr* 1985;4:171.
29. Willett WC, Sampson L, Browne ML, Stampfer MJ, Rosner B, Hennekens CH, Speizer FE. The use of a self-administered questionnaire to assess diet four years in the past. *Am J Epidemiol* 1988;127:188-99.
30. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semi-quantitative food frequency questionnaire. *Am J Epidemiol* 1985;122:51-65.
31. Fuchs C, Giovannucci E, Colditz G, Hunter D, Speizer F, Willett W. A prospective study of family history and the risk of colorectal cancer. *N Engl J Med* 1994;331:1669-74.
32. Stampfer MJ, Willett WC, Speizer FE, Dysert DC, Lipnick R, Rosner B, Hennekens CH. Test of the National Death Index. *Am J Epidemiol* 1984;119:837-9.
33. D'Agostino RB, Lee MLT, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Statistics in Medicine* 1990;9:1501-15.
34. Hosmer D, Lemeshow S. *Applied logistic regression*. New York: Wiley-Interscience, 1989.
35. Marshall R, Chisholm E. Hypothesis testing in the polychotomous logistic regression model with an application to detecting gastrointestinal cancer. *Statistics Med* 1985;4:337-44.
36. Glynn R, Rosner B. Methods to evaluate risks for composite endpoints and their individual components. *J Clin Epidemiol* 2004; in press.
37. Giovannucci E. Insulin and colon cancer. *Cancer Causes Control* 1995;6:164-79.
38. Terry PD, Miller AB, Rohan TE. Obesity and colorectal cancer risk in women. *Gut* 2002;51:191-4.
39. Konishi K, Fujii T, Boku N, Kato S, Koba I, Ohtsu A, Tajiri H, Ochiai A, Yoshida S. Clinicopathological differences between colonic and rectal carcinomas: are they based on the same mechanism of carcinogenesis? *Gut* 1999;45:818-21.
40. Servomaa K, Kiuru A, Kosma V-M, Hirvikoski P, Rytomaa T. p53 and K-ras gene mutations in carcinoma of the rectum among Finnish women. *Molec Pathol* 2000;53:24-30.
41. Terry PD, Miller AB, Rohan TE. Prospective cohort study of cigarette smoking and colorectal cancer risk in women. *Int J Cancer* 2002;99:480-3.