Assessment of Association between Fiber Intake and occurrence of Colorectal Cancer among 50-79 year old women, Washington State, 1981-1982.

AEPI 538D Earl Fields

This study was conducted to investigate the effect of various dietary factors on the occurrences of colorectal cancer. Currently colorectal cancer is the third most common cancer among both men and women in the United States. The study was intended to determine whether dietary factors and/or hormone is from a case-control study that attempted to determine if there was an association between some of these risk factors and colorectal cancer risk. The study was composed of cases were women aged 50 to 79, and identified within a few weeks after their diagnosis with colon or rectum cancer. Information about the subjects was obtained through from a population-based cancer registry in the Western Washington State region. The data was collected multiple sources including all hospitals, laboratories and radiation treatment facilities serving the area. The control subjects in the study were identified through a random digit telephone dialing, and were age matched to the study cases. The participants were contacted by mail and phone, and then the subjects were met for in-person interviews with trained, female interviewers.

The criteria for the cases and controls is indicated below:

<u>ELIGIBILITY CRITERIA FOR CASES</u>

- Female, aged 50-79 at diagnosis
- Diagnosed with cancer of the colon or rectum, but not anus: (ICDO1 153.0-154.1)
- Diagnosed between 11/1/1981 and 12/31/1982
- Histology: Adenocarcinoma
- Resident of a specific 3-county area including Seattle, Washington, at the time of diagnosis

- Able to speak English without assistance
- No previous cancer
- Had access to a residential telephone at the time of diagnosis – for example, NOT a resident of a nursing home with a central telephone

ELIGIBILITY CRITERIA FOR CONTROLS

- Female, aged 50-79 at the time the household was first contacted
- Never diagnosed with cancer prior to contact
- Resident of the same area as the cases, at the time the household was contacted
- Able to speak English without assistance
- Household of residence accessed and eligible through random digit dialing procedure, including:
- Accessed through a residential telephone number – for example, NOT a resident of a nursing home with a central telephone; NOT contacted initially at a work telephone number
- Matched to cases by age [i]

The variables you will be analyzing from the colorectal cancer dataset this semester are: CACO (case-control status), the OUTCOME variable; AGEDX (age at diagnosis or reference date); FAMCAN (family history of colorectal cancer; EDGP (education level); Monthly FIBER Intake will serve as the primary exposure variable; and BMI will serve as an additional potential confounder/ effect modifier.

For the agedx variable I used the levels defined by the course document to generate the categories. For the variables FIBER, and BMI I used the quadrennial to generate the levels. In each case the first or lowest level was used as the reference level.

The null hypothesis you will be testing: H_0 : The amount of fiber consumptions among those in the case control status 0 = the amount of fiber consumptions among those in the case control status 1.

Methods:

The data files were merged into a SAS data file and the data was cleaned by inspection, to remove values that appear to result to error in data entry. The 25th, 50th and 75th quadrennials

of the control group to create the cut points for the groups in this data analysis. These were used for all the exposure variables. The chi-sq values were used to initially describe the data and incorporated into table 1 showing the distribution of the data across the selected groups.

The effect of age was evaluated by adjusted the odd ratios of each of the variables of interest on the outcome variable. The same procedure was used to determine the effect of the primary exposure variable on the other variables looking at the OR and the associated heterogeneous p-value to look for evidence of a statistical difference. The data was then modeled using logistic regression and the method was compared to the results obtain in a crude stratified analysis to that obtained in a crude logistic regression. The data was modeled using logistic regression adjusted for and conditioned upon the age of the subject at the time of diagnosis.

The find the best model for the data set, logistic regression with backwards elimination was used to determine the most parsimonious model and the fully adjusted model. These models were analyzed for statistically significant p-values and confounding using the 10% rule and this produced the final model. Using the final model the effect of the age distribution size was evaluated by comparing the final model adjusted for age at 3 levels and at 6 levels of stratification.

Results:

Table 1: Based on the findings of the chi-square for independence, we find that there is a significant difference between the case and control categories for the variables AGE group, and Family History of cancer. All other variables fail to show a statistical difference between the case and control groups when evaluated independently.

Table 2: In determining if association between the cases of colorectal cancer and FIBER exists, the data showed that based on the crude values there is not a statistically significant association

between the variable FIBER and incidence of colorectal cancer based on the p-values. The evidence does suggest a dose response with a trend p value < 0.01.

Table 3: There is evidence of effect with the variables when controlling for BMI. There is evidence of association when controlling simultaneously for education level and family cancer history at two of the four levels (category 2 and 4). There is also evidence of effect modification when controlling simultaneously for education level and family cancer history at the third level, which did not show evidence of association.

Table 4: In the logistic regression analysis there was significance seen in some terms, age at diagnosis, family history of cancer, and education level completed. Using the 10% rule I did find evidence of confounding, at the FIBERCAT category 1 level only is based on the change in the OR values. This did not carry true to the other levels of the fiber categories. There was no evidence of confounding with the age-adjusted ratios for any of the other variables.

Table 5: In using logistic regression analysis to determine if there was interaction between the different times used in this modeling, none of the interaction terms showed significance at the p < 0.05 level, the effected of FIBER intake was evaluated, while controlling for different levels of Age at diagnosis, family history of cancer, and education level completed. However using the 10% rule I do not find evidence of effect modification, this is based on the change in the OR values as the interactions are included/removed.

Table 6: Backwards regression was used to eliminate all terms which do not illustrate a statistically significance. BMI and Education level completed were removed from the model as a result to generate the most parsimonious model.

Table 7: Using logistic regression at the p-value of < 0.05 and the 10 % rule the final model was produced, containing the variables Fiber Intake, Education level complete, Family Cancer

history, and BMI.

Table 8: The results for our exposure of interest Fiber Consumption was not varied significantly between the unconditional logistic regression model, where we adjusted for age, and the conditional logistic regression model in which Fiber Content was conditioned upon age. Furthermore, when age was categorized as a 6-level variable, with 5 year age groups, the conditional logistic regression results did not differ significantly from those obtained from the regression model in which AGE was categorized as a 3 level variable, with 10 year age groups.

Conclusion:

Based on the finding outline in the analysis of this study, there appears to be no modification of the effect of the occurrence of colorectal cancer as a direct result of monthly fiber intake. However, when this data is considered in the presence of other factors including the subjects age at diagnosis, highest education level completed and family history of cancer, there does appear to be a confounding of the relation between fiber intake at the highest and lowest levels of consumption.

[i] Course document from blackboard: "Description_COLORECTAL_CANCER_data_files_final.doc"

Table 1. Characteristics of subjets from Colon Cancer assessment (Cases and age matched Controls)

Characteristic	Cases		Controls			
	n=192	%	n=295	%	X^2 (d.f.)*	*p-value*
AGE IN YEARS AT TIME OF DIAGNOSIS						
50-59	66	*13.6%*	69	*14.2%*		
60-69	117	*24.0%*	72	*14.8%*		
70-79	112	*23.9%*	51	*10.5%*	12.4 (2)	*0.0021*
INDEX OF CALCIUM INTAKE						
< 117	76	*15.6%*	47	*9.7%*		
117 - 179	95	*19.5%*	48	*9.9%*		
180 - 249	68	*14.0%*	49	*10.1%*		
> 250	56	*11.5%*	48	*10.0%*	4.39 (3)	*0.2224*
SERVINGS OF CHEESE CONSUMED PER MONTH						
<4	56	*11.5%*	24	*4.9%*		
4-7	51	*10.5%*	42	*8.6%*		
8-16	114	*23.4%*	68	*14.0%*		
>16	74	*15.2%*	58	*11.9%*	5.70 (3)	*0.1274*
missing	2					
CUPS OF COFFEE CONSUMED PER MONTH						
<59	77	*15.8%*	46	*15.8%*		
60-90	65	*13.4%*	37	*7.6%*		
90-151	88	*18.1%*	49	*10.1%*		
>151	65	*13.4%*	60	*12.3%*	5.24 (3)	*0.155*
INDEX OF FIBER INTAKE						
>3.70	111	*11.7%*	50	*10.3%*		
3.70-4.87	64	*13.1%*	45	*9.2%*		
4.88-6.40	63	*12.9%*	49	*10.1%*		
>6.40	57	*11.7%*	48	*9.9%*	7.48 (3)	0.0581
HIGHEST EDUCATION COMPLETED						
No High School	93	*19.1%*	37	*7.6%*		
Completed High School	126	*25.9%*	79	*16.2%*		
Some College	76	*15.6%*	76	*15.6%*		
PARENT OR SIBLING WITH CANCER HISTORY						
Yes	50	*10.3%*	19	*3.9%*		
No	245	*50.5%*	171	*35.3%*	4.56 (1)	*0.0327*
BMI						
0-20.0	32	*6.6%*	26	*5.4%*		
20.0-24.9	157	*32.6%*	106	*22.0%*		
25.0-29.9	69	*14.3%*	40	*8.3%*		
> 30	32	*6.6%*	20	*4.2%*	1.11 (3)	*0.7738*
missing			5			

* Chi-square d.f. = degrees of freedom

Table 2. Unadjusted and age-adjusted associations of outcome status based on variables fiber intake, cheese consump and calcium intake stratified by age group.

Characteristic	Crude			Age-Adjusted*			
	Odds Ratio	95% C.I.	trend ¥ p-value	Odds Ratio	95% C.I.†	heterog. p-value ^	trend ¥ p-value
AGE IN YEARS AT TIME OF DIAGNOSIS							
50-59	1	--		--			
60-69	1.70	(1.09-2.66)		---			
70-79	2.30	(1.43-3.68)		---			
INDEX OF FIBER INTAKE - PRIMARY			0.01				0.01
>3.70 (Reference)	1	---		1	---		
3.70-4.87	0.64	(0.356-1.063)		0.61	(0.364-1.03)	0.76	
4.88-6.40	0.58	(0.351-0.956)		0.57	(0.338-0.950)	0.64	
>6.40	1.87	(1.12-3.11)		2.07	(1.22-3.51)	0.20	
FAMILY HISTORY OF CANCER			0.03				0.042
No (Reference)	1	---		1	---		
Yes	1.84	(1.05-3.23)		1.78	(1.01-3.14)	0.74	
BMI			0.38				0.37
0-20.0 (Reference)	1	---		1	---		
20.0-24.9	1.20	(0.68-2.13)		1.26	(0.681-2.333)	0.28	
25.0-29.9	1.40	(0.73-2.68)		1.32	(0.662-2.629)	0.27	
> 30	0.77	(0.359-1.65)		0.72	(0.3233-1.611)	0.27	
HIGHEST EDUCATION LEVEL COMPLETED			0.0002				0.001
NO HIGH SCHOOL (Reference)	1.00	--		-			
COMPLETED HIGH SCHOOL	1.58	(0.982-2.53)		1.47	(0.904-2.38)	0.92	
SOME COLLEGE	2.51	(1.530-4.13)		2.25	(1.356-3.734)	0.81	

* Age-adjusted by categories: 50-59, 60-69, 70-79

† C.I. Confidence interval

^ Breslow-Day test for heterogeneity of odds ratios

¥ Mantel-Haenszel chi-square

Table 3. Association of Monthly Fiber Intake with Cancer Case Control status, adjusted for other potential confounders

Adjusted For *	Monthly Fiber Intake	Odds Ratio	*95% C.I.* †	*heterog. p-value* ^	*Trend p-value* ¶
Unadjusted (crude)	>3.70 (Reference)	1.00	--		*0.0108*
	3.70-4.87	0.64	*(0.39-1.06)*		
	4.88-6.40	0.58	*(0.35-0.96)*		
	>6.40	0.53	*(0.32-0.89)*		
Age	>3.70 (Reference)	1.00	--		*0.01*
	3.70-4.87	0.61	(0.364-1.031)	0.76	
	4.88-6.40	0.57	(0.338-0.950)	0.64	
	>6.40	0.49	(0.29-0.82)	0.20	
Family Cancer History	>3.70 (Reference)	1.00	--		*0.01*
	3.70-4.87	0.60	*(0.360-1.01)*	*0.69*	
	4.88-6.40	0.55	*(0.334-0.919)*	*0.86*	
	>6.40	0.53	*(0.31-0.88)*	*0.47*	
Highest Education level Completed	>3.70 (Reference)	1.00	--		*0.02*
	3.70-4.87	0.58	*(0.34-0.98)*	*0.93*	
	4.88-6.40	0.57	*(0.34-0.95)*	*0.69*	
	>6.40	0.32	*(0.323-0.904)*	*0.87*	
BMI	>3.70 (Reference)	1.00	--		*0.01*
	3.70-4.87	0.65	*(0.39-1.07)*	*0.03*	
	4.88-6.40	0.56	*(0.34-0.93)*	*0.02*	
	>6.40	0.54	*(0.32-0.90)*	*0.30*	
Family Cancer History * Education					
	>3.70 (Reference)	1.00	--		*0.01*
	3.70-4.87	0.62	*(0.228-1.712)*	0.03	
	4.88-6.40	0.42	*(0.177-0.98)*	0.53	
	>6.40	0.43	*(0.193-0.970)*	0.04	
Family Cancer History * Education					
No High School	>3.70 (Reference)	1.00	--		*0.35*
	3.70-4.87	0.52	*(0.182-1.462)*		
	4.88-6.40	0.63	*(0.21-1.891)*		
	>6.40	0.52	*(0.163-1.631)*		
Completed High School	>3.70 (Reference)	1.00	--		*0.06*
	3.70-4.87	0.57	*(0.2658-1.200)*		
	4.88-6.40	0.70	*(0.311-1.56)*		
	>6.40	0.47	*(0.213-1.041)*		
Some College	>3.70 (Reference)	1.00	--		*0.18*
	3.70-4.87	0.68	*(0.249-1.864)*		
	4.88-6.40	0.42	*(0.181-0.991)*		
	>6.40	0.64	*(0.279-1.487)*		

* For categories of adjustment variables, see Table 2

† C.I. Confidence interval

^ Breslow-Day test for heterogeneity of odds ratios

¶ Test for Trend (Mantel extension)

Table 4. Unadjusted and adjusted odds ratios of various characteristics with colorectal cancer among 50-79 year old women, Washington State, 1981-1982

Characteristic		Stratified Analysis Crude		Logistic regression analysis Crude			Logistic regression analysis Adjusted*		
		Odds Ratio	*95% C.I.* †	Odds Ratio	95% C.I.†	p-value‡	Odds Ratio	95% C.I.†	p-value‡
BMI						*0.774*			*0.881*
Low (< 20)		*1.00*	---	1.00	--		1	---	
Med (20-24.9)	BMICAT1	*1.20*	*(0.68-2.13)*	1.20	*(0.68-2.14)*		1.18	*(0.65-2.17)*	
Mod (25-29.9)	BMICAT2	*1.40*	*(0.73-2.68)*	1.40	*(0.73-2.68)*		1.30	*(0.66-2.60)*	
High (> =30)	BMICAT3	*1.30*	*(0.61-2.78)*	1.30	*(0.61-2.78)*		1.31	*(0.58-2.92)*	

CRUDE MODEL: logit (P(D=1 | BMICAT)) = b0 + b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3

ADJUSTED MODEL: logit (P(D=1 | BMICAT, EDGP, FIBERCAT, FAMCAN, AGECAT)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*EDGP1 + b5*EDGP2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) +b11*FAMCAN b11*FAMCAN

Education						*0.001*			*0.003*
No high school education		*1*	---	1	---		1	---	
Completed HS	EDGP1	0.63	(0.40-1.02)	0.64	*(0.40-1.02)*		0.68	*(0.42-1.12)*	
College or greater	EDGP2	0.40	(0.24-0.65)	0.40	*(0.24-0.65)*		0.41	*(0.24-0.70)*	

CRUDE MODEL: logit (P(D=1 | EDGP)) = b0 + b1*EDGP1 + b2*EDGP2

ADJUSTED MODEL: logit (P(D=1 | BMICAT, EDGP, FIBERCAT, FAMCAN, AGECAT)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*EDGP1 + b5*EDGP2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) + b11*FAMCAN b11*FAMCAN

AGE IN YEARS AT TIME OF DIAGNOSIS						*0.002*			*0.004*
50-59		1	--	1	---		1	---	
60-69	AGECAT1	1.70	(1.09-2.66)	1.70	(1.09-2.66)		1.73	*(1.08-2.77)*	
70-79	AGECAT2	2.30	(1.43-3.68)	2.30	(1.43-3.68)		2.29	*(1.39-3.78)*	

CRUDE MODEL: logit (P(D=1 | AGECAT)) = b_0 + b_1*AGECAT1 + b_2*AGECAT2

ADJUSTED MODEL: logit (P(D=1 | BMICAT, EDGP, FIBERCAT, FAMCAN, AGECAT)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*EDGP1 + b5*EDGP2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) + b11*FAMCAN

Table 5. Adjusted* odds ratios for FIBER INTAKE and colorectal cancer with various effect modifiers, among 50-79 year old women, Washington State, 1981-1982

Interaction with	FIBER INDEX Odds Ratios					
	Moderate (3.70-4.87):Low (>3.70)		Med-high (4.88-6.40):Low (>3.70)		High (>6.40):Low (>3.70)	
	Odds Ratio	95% C.I.†	Odds Ratio	95% C.I.†	Odds Ratio	95% C.I.†
No Interaction	0.52	*(0.31-0.89)*	0.52	*(0.31-0.88)*	0.47	*(0.27-0.80)*
ADJUSTED MODEL: logit (P(D=1 | FIBERCAT, BMICAT, EDGP, FAMCAN, AGECAT)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*EDGP1 + b5*EDGP2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) +b11*FAMCAN b11*FAMCAN						
Education						
Not a high school graduate - Ref	0.50	*(0.173-1.445)*	0.58	*(0.188-1.788)*	0.50	*(0.154-1.597)*
High School Graduate - Educ1	0.47	*(0.215-1.027)*	0.64	*(0.279-1.476)*	0.38	*(0.167-0.856)*
College or greater - Educ2	0.69	*(0.242-1.953)*	0.38	*(0.158-0.921)*	0.58	*(0.239-1.427)*
homogeneity p-value =0.8540‡						
ADJUSTED MODEL: logit (P(D=1 | BMI, Educ AGECAT, Age)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*Educ1 + b5*Educ2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) + (b11*Age1 + b12*Age2) + + (b_{13}*FIBERCAT1*Educ1) + (b_{14}*FIBERCAT1*Educ2) + (b_{15}*FIBERCAT2*Educ1) + (b_{16}*FIBERCAT2*Educ2) + (b_{17}*FIBERCAT3*Educ1) + (b_{18}*FIBERCAT3*Educ2)						
BMI						
LOW (1st quartile) (>20) - Reference	1.49	*(0.282-7.824)*	0.26	*(0.046-1.438)*	0.57	*(0.147-2.188)*
MED-LOW 2nd quartile (20-24.9) - BMICAT1	0.25	*(.124-.518)*	0.37	*(0.179-0.766)*	0.33	*(0.151-0.699)*
MODERATE 3rd quartile (25-29.9) - BMICAT2	1.15	*(0.331-4.020)*	1.26	*(0.443-3.592)*	0.89	*(0.289-2.745)*
HIGH 4th quartile (< 30) - BMICAT3	1.94	*(0.372-10.087)*	0.66	*(0.119-3.615)*	0.91	*(0.209-3.943)*
homogeneity p-value = 0.1633 ‡						
ADJUSTED MODEL: logit (P(D=1 | FIBERCAT, BMICAT, EDGP, FAMCAN, AGECAT)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*Educ1 + b5*Educ2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) + (b11*Age1 + b12*Age2) + + (b13*BMICAT1*FIBERCAT1) + (b14*BMICAT2*FIBERCAT1) + (b15*BMICAT3*FIBERCAT1) + (b16*BMICAT1*FIBERCAT2) + (b17*BMICAT2*FIBERCAT2) + (b18*BMICAT3*FIBERCAT2) + (b19*BMICAT1*FIBERCAT3) + (b20*BMICAT2*FIBERCAT3) + (b21*BMICAT3*FIBERCAT3)						
AGE IN YEARS AT TIME OF DIAGNOSIS						
50-59 Reference	0.76	*(0.287-2.010)*	0.51	*(0.202-1.287)*	1.07	*(0.387-2.936)*
60-69 AGECAT1	0.46	*(0.196-1.063)*	0.62	*(0.262-1.461)*	0.28	*(0.109-0.697)*
70-79 AGECAT2	0.40	*(0.145-1.096)*	0.37	*(0.134-1.037)*	0.40	*(0.162-1.001)*
homogeneity p-value = 0.430 ‡						
ADJUSTED MODEL: logit (P(D=1 | FIBERCAT, BMICAT, EDGP, FAMCAN, AGECAT)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*Educ1 + b5*Educ2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) + (b11*Age1 + b12*Age2) + + (b13*AGECAT1*FIBERCAT1) + (b14*AGECAT1*FIBERCAT2) + (b15*AGECAT1*FIBERCAT3) + (b16*AGECAT2*FIBERCAT1) + (b17*AGECAT2*FIBERCAT2) + (b18*AGECAT2*FIBERCAT3)						
FAMILY HISTORY OF CANCER						
No	0.51	*(0.284-0.898)*	0.55	*(0.315-0.969)*	0.45	*(0.254-0.788)*
Yes	0.63	*(0.131-2.982)*	0.38	*(0.084-1.723)*	0.71	*(0.120-4.204)*
homogeneity p-value = 0.6369‡						
ADJUSTED MODEL: logit (P(D=1 | FIBERCAT, BMICAT, EDGP, FAMCAN, AGECAT)) = b0 + (b1*BMICAT1 + b2*BMICAT2 + b3*BMICAT3) + + (b4*Educ1 + b5*Educ2) + (b6*AGECAT1 + b7*AGECAT2)+ (b8*FIBERCAT1 + b9*FIBERCAT2 + b10*FIBERCAT3) + (b11*Age1 + b12*Age2) + + (b13*FAMCAN*FIBERCAT1) + (b14*FAMCAN*FIBERCAT2) + (b15*FAMCAN*FIBERCAT3)						

* Models include BMI, age at diagnosis/first contact, education, and fiber & family cancer history.
† C.I. Confidence interval
‡ Homogeneity p-value is from the Wald chi-square test for the significance of the combined cross-product terms

Table 6: Backwards Elimination of Crossproduct and Main Effects Terms

For the situation where there is NO INTERACTION

Backwards Elimination of Crossproduct Terms

Initial Model: Logit (P(D=1|fibercat,bmicat,agecat,famcan) = b0 + [b1-b3]*[fibercat1-3] + [b4-b6]*[bmicat1-3] + [b7-b8]*[agecat1-2] + b9*famcan + [b10-b12]*[edgp1-3] + [b13-b21]*[fibercat1-3]*[bmicat1-3] + [b22-b27]*[fibercat1-3]*[agecat1-2] + [b28-b30]*[fibercat1-3]*famcan + [b31-b33]*[

	Cross-Product Term	Removed or retained p-value
Step 1:	FIBERCAT*edgp	0.9152 (removed)
Step 2:	FIBERCAT*famcan	0.6985 (removed)
Step 3:	FIBERCAT*AGECAT	0.4097 (removed)
Step 4:	FIBERCAT*BMICAT	0.1633 (removed)

Fully Adjusted Model: Logit (P(D=1|bmicat,fibercat,edgp,agecat,famcan) = b0 + [b1-b3]*[fibercat1-3] + [b4-b6]*[bmicat1-3] + [b7-b8]*[agecat1-2] + b9*famcan + [b10-b12]*[edgp1-3]

Backwards Elimination of Main Effects Terms

	Fully Adjusted Model
Fiber Consumption	
low (LESS THAN 3.70)	**1**
medium (3.70 to 4.87)	**0.522**
moderate (4.87 to 6.40)	**0.520**
high (MORE THAN 6.40)	**0.469**

Step1:

Main Effects Term	Removed or retained p-value
BMICAT	**0.8808 (removed)**

	Step 1 Model
Fiber Consumption	
low (LESS THAN 3.70)	**1**
medium (3.70 to 4.87)	**0.524**
moderate (4.87 to 6.40)	**0.530**
high (MORE THAN 6.40)	**0.469**

Step 2:

Main Effects Term	Removed or retained p-value
famcan	**0.0326 (retained)**

	Step 2 Model
Fiber Consumption	
low (LESS THAN 3.70)	**1**
medium (3.70 to 4.87)	**0.548**
moderate (4.87 to 6.40)	**0.551**
high (MORE THAN 6.40)	**0.472**

Step 3:

Main Effects Term	Removed or retained p-value
edgp	**0.0039 (retained)**

	Step 3 Model
Not completed High School	**1**
Completed H.S.	
Some College	
Fiber Consumption	
low (LESS THAN 3.70)	
medium (3.70 to 4.87)	
moderate (4.87 to 6.40)	**1**
high (MORE THAN 6.40)	**0.607**
	0.552
	0.468

Table 7: Logistic regression summary: fully adjusted, most parsimonious and final models. Fiber consumption and colorectal cancer risk, among 50-79 year old women, Washington State, 1981-1982.

For the Situation where there is NO INTERACTION

		Fully Adjusted Model			Most Parsimonious Model			Final Model A		
		OR*	**95% C.I.†**	***Wald p-value ^***	**OR**	**95% C.I.†**	***Wald p-value ^***	**OR**	**95% C.I.†**	***Wald p-value ^***
Fiber Index				0.0155			0.0048			0.0197
low (LESS THAN 3.70)		**1**	--		1	--		1.0	--	
medium (3.70 to 4.87)		**0.522**	(0.305-0.892)		0.61	(0.36-1.02)		0.519	(0.304- 0.885)	
moderate (4.87 to 6.40)		**0.520**	(0.307-0.882)		0.55	(0.330-0.924)		0.536	(0.318- 0.904)	
high (MORE THAN 6.40)		**0.469**	(0.274-0.804)		0.47	(0.269-7.93)		0.487	(0.286- 0.830)	
		trend‡ p =0.0044			trend‡ p = .015			trend‡ p =0.0069		
BMI				0.8808						
low (LESS THAN 20)		**1**	--							
medium (20 to 24.9)		**1.2**	(0.646-2.169)							
moderate (25 to 29.9)		**1.3**	(0.656-2.595)							
high (MORE THAN 30)		**1.3**	(0.584-2.918)							
		trend‡ p =0.4627								
Edu				0.0034						0.0026
Not completed High School		**1**	--					**1.0**	--	
Completed H.S.		**0.684**	(0.417-1.122)					**0.66**	(0.403- 1.079)	
Some College		**0.412**	(0.244-0.697)					**0.405**	0.241- 0.681)	
		trend‡ p =0.0008						trend‡ p =0.0006		
Age				0.0044			0.0005			0.0075
50-59		**1**	--		**1**	--		**1**	--	
60-69	AGECAT1	**1.73**	(1.077- 2.772)		**1.81**	(1.14-2.87)		**1.674**	(1.049- 2.672)	
70-79	AGECAT2	**2.29**	(1.385- 3.778)		**2.38**	(1.47-3.87)		**2.173**	(1.323 -3.569)	
		trend‡ p = 0.0013			trend‡ p = 0.0013			trend‡ p =0.0022		
Family history of cancer				0.0352			0.0338			0.0183
No (Reference)		**1**	--		1			**1**	--	
Yes		**1.9**	(1.05-3.384)		1.9	(1.055-3.404)		**2.01**	(1.126- 3.590)	
		trend‡ p =0.0352			trend‡ p = 0.0338			trend‡ p = 0.0013		
		HL GOF test♦:	0.8808		HL GOF test♦:	0.8153		HL GOF test♦: 0.7026		

All logistic regression models included age (50-59, 60-69, 70-79) as well as the other indicated variables

*OR=Odds Ratio

†95% C.I. = 95% Confidence Interval

^Wald p-value = chunk test for overall significance of variable

‡test for trend: significance of beta for an ordinal variable

♦: Hosmer-Lemeshow goodness of fit test (pvalue)

Table 8: Fiber consumption and colorectal cancer risk, among 50-79 year old women, Washington State, 1981-1982.
Logistic regression summary: Comparison of unconditional model, and models conditioned on age.

	Final Model Unconditional Logistic Adjusted for age (50-59, 60-69, 70-79)			Final Model Conditional Logistic Conditioned on age (50-59, 60-69, 70-79)			Final Model Conditional Logistic Conditioned on age (50-54, 55-59, 60-64, 65-69, 70-74, 75-79)		
	OR*	**95% C.I.†**	***Wald p-value ^***	**OR***	**95% C.I.†**	***Wald p-value ^***	**OR***	**95% C.I.†**	***Wald p-value ^***
Fiber Index			0.0155			0.016			0.0194
low (LESS THAN 3.70)	**1**	--		1.0	--		1.0	--	
medium (3.70 to 4.87)	**0.522**	(0.305-0.892)		0.524	(0.307-0.894)		0.531	(0.311-0.906)	
moderate (4.87 to 6.40)	**0.520**	(0.307-0.882)		0.523	(0.307-0.882)		0.53	(0.313-0.897)	
high (MORE THAN 6.40)	**0.469**	(0.274-0.804)		0.471	(0.274-0.804)		0.478	(0.279-0.819)	
	trend‡ p =0.0044			trend‡ p = 0.0046			trend‡ p = 0.0055		
BMI			0.8808			0.8818			0.9097
low (LESS THAN 20)	**1**	--		1.0	--		1.0	--	
medium (20 to 24.9)	**1.18**	(0.646-2.169)		1.18	(0.646-2.162)		1.19	(0.648-2.170)	
moderate (25 to 29.9)	**1.30**	(0.656-2.595)		1.30	(0.656-2.584)		1.28	(0.646-2.545)	
high (MORE THAN 30)	**1.31**	(0.584-2.918)		1.30	(0.584-2.906)		1.25	(0.557-2.806)	
	trend‡ p =0.4627			trend‡ p =0.4639			trend‡ p =0.5450		
Edu			0.0034			0.0036			0.0029
Not completed High School	**1**	--		1.0	--		1.0	--	
Completed H.S.	**0.684**	(0.417-1.122)		0.685	(0.418-1.123)		0.684	(0.416-1.122)	
Some College	**0.412**	(0.244-0.697)		0.415	(0.246-0.700)		0.407	(0.241-0.689)	
	trend‡ p =0.0008			trend‡ p =0.0009			trend‡ p =0.0007		
Family history of cancer			0.0352			0.0359			0.0432
No (Reference)	**1**	--		1	--		1	--	
Yes	**1.9**	(1.05-3.384)		1.9	(1.042-3.364)		1.835	(1.019-3.307)	
	trend‡ p =0.0352								
	HL GOF test♯: 0.74	0.8808		(no HL GOF test)			(no HL GOF test)		

*OR=Odds Ratio
†95% C.I = 95% Confidence Interval
^Wald p-value = chunk test for overall significance of variable in the model
‡test for trend: significance of beta for an ordinal variable
♯: Hosmer-Lemeshow goodness of fit test