

# Is principal components analysis necessary to characterise dietary behaviour in studies of diet and disease?

Susan E McCann\*, John Weiner, Saxon Graham and Jo L Freudenheim  
Department of Social and Preventive Medicine, 270 Farber Hall, University at Buffalo, Buffalo,  
NY 14214, USA

Submitted 26 September 2000; Accepted 1 January 2001

## Abstract

*Objective:* To assess the relative ability of principal components analysis (PCA)-derived dietary patterns to correctly identify cases and controls compared with other methods of characterising food intake.

*Subjects:* Participants in this study were 232 endometrial cancer cases and 639 controls from the Western New York Diet Study, 1986–1991, frequency-matched to cases on age and county of residence.

*Design:* Usual intake in the year preceding interview of 190 foods and beverages was collected during a personal interview using a detailed food-frequency questionnaire. Principal components analysis identified two major dietary patterns which we labelled 'healthy' and 'high fat'. Classification on disease status was assessed with separate discriminant analyses (DAs) for four different characterisation schemes: stepwise DA of 168 food items to identify the subset of foods that best discriminated between cases and controls; foods associated with each PCA-derived dietary pattern; fruits and vegetables (47 items); and stepwise DA of USDA-defined food groups (fresh fruit, canned/frozen fruit, raw vegetables, cooked vegetables, red meat, poultry, fish and seafood, processed meats, snacks and sweets, grain products, dairy, and fats).

*Results:* In general, classification of disease status was somewhat better among cases (54.7% to 67.7%) than controls (54.0% to 63.1%). Correct classification was highest for fruits and vegetables (67.7% and 62.9%, respectively) but comparable to that of the other schemes (49.5% to 66.8%).

*Conclusions:* Our results suggest that the use of principal components analysis to characterise dietary behaviour may not provide substantial advantages over more commonly used, less sophisticated methods of characterising diet.

**Keywords**  
Dietary patterns  
Food patterns  
Statistical methods  
Misclassification

The information obtained with food-frequency questionnaires (FFQs) has been summarised in a variety of ways in studies of diet and disease. The most common use of food-frequency information is calculation of the intake of nutrients believed to be related to risk of disease. While this approach has utility, examination of single nutrients neglects the non-independence of nutrient intake. Individuals do not exist on single nutrients taken in isolation, but on combinations of specific foods often eaten together. Thus, an analysis limited to correlations of specific nutrients to risk of a given disease is inadequate. Examination of foods and food groups partly addresses the issue of dependence. It has the added advantage of being amenable to the formation of subsequent public health recommendations.

Principal components analysis (PCA) is a statistical technique that converts a large number of related

variables into a smaller set of 'factors' intended to describe specific patterns of behaviour. Each factor is composed of a weighted set of the original variables. The weights (factor loadings) or coefficients are selected to render each factor independent of the others and to sequentially explain the largest amount of the total variance possible. This analytical model was introduced to nutritional epidemiology as an attempt to capture the multidimensionality of diet while reducing the apparent complexity<sup>1–5</sup>. The few diet studies that have used PCA-derived dietary patterns to estimate disease risk have been careful to establish that the patterns were correlated with specific nutrient intake and described an interpretable pattern of dietary behaviour<sup>6–9</sup>. Although these studies provided evidence that PCA patterns do describe diet, and can be used to estimate risk of disease, they have not addressed whether they are superior to other commonly used, but less

statistically sophisticated, methods of characterising food use (e.g. expert-selected foods or food groups).

The characterisation of dietary intake is important in linking diet to chronic disease risk. The characterisation process has been an important issue in dietary assessment methodology<sup>10</sup>. To estimate disease risk accurately, the dietary intake of subjects must be appropriately described to limit misclassification of exposure and provide the best discrimination of subjects as to disease. The estimation of disease risk associated with diet depends on the degree to which the dietary behaviour of a subject with disease differs from that of a subject without disease, as well as the extent to which the method of characterisation captures these differences. If the method used to characterise dietary intake is inaccurate with regard to measurement of the characteristics that distinguish diseased from non-diseased individuals, the diet of the diseased subjects may appear more like the diet of the controls, thus reducing our ability to identify dietary risk factors.

In this study, we were interested in how well the foods associated with dietary patterns identified through PCA could correctly distinguish subjects as cases or controls compared with simpler methods of characterising diet. To accomplish this goal, we used linear discriminant analysis (LDA) to predict correct classification of endometrial cancer cases versus controls by different methods of describing dietary intake (PCA dietary patterns, food groups, single foods). Linear discriminant analysis is a statistical technique similar to regression analysis except that the dependent variable is categorical rather than continuous. Whereas the goal of regression is to predict the value of a variable of interest based on a set of predictor variables, in LDA the goal is to predict class membership of an individual observation based on a set of predictor variables<sup>11</sup>. In the present analyses, four major methods of describing food use were assessed for their ability to predict case-control status. The first involved, simply, 168 individual foods and beverages from the food-frequency questionnaire used to obtain dietary data. The second involved foods most important in describing three different principal components analysis models. The third involved all 47 of the fruits and vegetables from the FFQ. The fourth involved 12 USDA-defined food groups.

## Materials and methods

The present study utilised data from a series of case-control studies of diet and cancer of the breast, endometrium, ovary and prostate in western New York, the methods for which have been described in detail elsewhere<sup>12</sup>. The study protocol for the Western New York Diet Study was approved by the Institutional Review Board of the State University of New York at Buffalo, and informed consent was obtained from all subjects. For each

cancer site, incident cases were identified by trained nurse case-finders. Controls were frequency-matched to each case on gender, age and county of residence. Data were collected during extensive interviewer-administered interviews for all participants in the study between 1986 and 1991. Diet in the year two years before the interview was queried by a series of detailed questions regarding usual frequency and quantity of intake of 190 foods and beverages. Portion size for each food was assessed using food models, and additional questions were included regarding seasonal variation of use and food preparation methods.

Dietary patterns were identified using dietary data from the entire Western New York Diet Study female control series ( $n = 863$ ). Principal components analysis may be affected by the number of items included in the determination of factors. Therefore, we had previously assessed the effect of using three separate methods to re-code weekly frequency of food use into smaller numbers of items for analysis prior to principal components analysis (McCann SE, Marshall JR, Brasure JR, Graham S, Freudenheim JL. Analysis of patterns of food intake in nutritional epidemiology: food classification in principal components analysis and the subsequent impact on estimates for endometrial cancer. *Public Health Nutr.*, accepted for publication). The results of those analyses indicated that the number of foods used in the PCA model did affect the amount of variance explained, but not the number or character of the patterns identified, as reducing the number of variables in the PCA model provided relatively comparable patterns that could be interpreted. We did, however, find slight differences in risk estimation associated with the dietary patterns depending on the level of detail included in the PCA. Therefore, the present analyses include the three sets of dietary patterns as identified in the previous analyses. A brief description of these methods follows.

For the first PCA, the original 190 foods on the FFQ were reduced slightly by averaging seasonal use and combining raw and cooked forms of the same food for a total of 168 individual foods and beverages. In the second PCA, the 168 individual food items were classified into categories based on nutrient content and usage into 56 detailed food groups (calculated as the sum of the weekly frequency of use of the foods or beverages assigned to each group) and individual foods and beverages (salad vegetables, cruciferous vegetables, high-carotenoid vegetables, other vegetables, fresh fruit, canned or frozen fruit, dried fruit, fruit juices, high-fat dairy products, low-fat dairy products, refined grains, whole grains, snacks, desserts, candy, nuts, red meat, pork, processed meats, fish, poultry, soup, and fast foods). Several foods and beverages were not categorised as they were thought to reflect specific dietary behaviours (coleslaw, pickles, mashed potatoes, baked potatoes, fried potatoes, other potatoes, baked beans, tomato juice, margarine, butter,

mayonnaise and miracle whip, salad dressing, reduced-calorie mayonnaise and salad dressing, yoghurt, eggs, jams or jellies, liver, macaroni and cheese, spaghetti or lasagna, pizza, tacos, chilli, pot pies, and non-dairy beverages). Finally, in the third PCA, the 56 groups and food items were further collapsed into 36 broader groups (calculated as described for the second PCA) based on USDA food classification definitions and nutrient content (vegetables, fruit, regular fats, reduced-calorie fats, high-fat dairy, low-fat dairy (including yoghurt), eggs, refined grains, whole grains, snacks, desserts, jams or jellies, candy, nuts, meat, processed meats, fish, poultry, and fast foods). Macaroni and cheese, spaghetti or lasagna, pizza, tacos, chilli, pot pies, soup, and non-dairy beverages were left as separate items.

For each method, dietary patterns were identified by principal components analysis in SPSS for Windows, version 8. All variables were expressed as standardised weekly frequency of use. Factors were rotated with an orthogonal (varimax) rotation to improve interpretability and minimise the correlation between the factors. The number of factors retained from each PCA was determined by the amount of variance explained by each factor, and by factor interpretability. Labelling of the factors was primarily descriptive and based on our interpretation of the pattern structures. Each PCA identified two interpretable factors that we labelled 'healthy' and 'high fat'. The patterns and the foods associated with each pattern are identified in Table 1.

To assess discrimination of disease status by different diet characterisation methods, we used dietary data from 232 women with incident, histologically confirmed endometrial cancer and 639 controls from the Western New York Diet Study. The different methods of characterising diet involved using foods from four different food-use characterisation schemes as shown in Table 1 and described below. The classification of cases and controls using the foods was performed using a series of linear discriminant analyses (LDAs).

For the first LDA, the original 190 items were reduced slightly by averaging frequencies of seasonal and non-seasonal items and combining frequencies of raw and cooked items for a total of 168 single food items entered into a stepwise selection procedure to obtain the set of foods most predictive of case-control membership. As classification of foods into groups may be somewhat subjective, we utilised the stepwise procedure to represent an objective method of choosing a subset of food items that would provide the highest percentage of cases and controls correctly classified. A second analysis used the foods identified for each PCA pattern in Table 1 to predict case-control membership wherein the food items most strongly associated with each pattern (factor loadings  $\geq 0.30$  or  $\leq -0.20$ ) were entered as a block into separate discriminant analyses. The third analysis involved 47 individual fruits and vegetables as an arbitrary

set of foods representing a 'healthy' diet. The fourth analysis included 12 food groups classified by USDA classification definitions and a stepwise procedure to determine the food groups most predictive of case-control membership. All variables were expressed as weekly frequency of use. The percentage correct classification of subjects for linear discriminant analysis of each food use characterisation method was then compared for differences.

## Results

The results of the discriminant analyses for the four methods of characterising diet are shown in Table 2. The first analysis identified the following foods as maximally separating cases and controls: saltines, corn, spinach, butter, olives, soup, bananas, tacos, beets, pancakes, and coffee. These 11 foods correctly classified 64.7% of subjects as cases and 63.1% as controls.

The second analysis showed different correct classifications for the patterns identified in the PCA. For both the 'healthy' and the 'high fat' patterns, correct classification of subjects as to case or control status decreased as the number of items involved in the PCA and the grouping of the foods became less detailed. This effect was slightly more evident for classification of cases by the 'healthy' patterns, with correct classification decreasing from 65.5% to 54.7% (168-item PCA to 36-item PCA) compared with a decrease from 66.8% to 59.9% for the 'high fat' patterns (168-item PCA to 36-item PCA).

Finally, discrimination of case-control status by more commonly used methods of characterising food use was examined. Compared with the stepwise discriminant analysis of the 168 individual items, use of an arbitrary group of foods (47 fruits and vegetables) produced slightly higher correct classification of cases (67.7% versus 64.7%), but not of controls (62.9% versus 63.1%). When food groups defined using USDA definitions were entered into a stepwise discriminant analysis, the highest correct classification of cases and controls was obtained from three groups (non-raw fruits, cooked vegetables, and grains), although classification was, in general, lower than that obtained using the other methods of characterising intake (57.3% and 59.2%, cases and controls, respectively).

## Discussion

The appropriate characterisation of dietary behaviour for estimation of disease risk has been the subject of numerous methodological inquiries. Because dietary intake is multidimensional and often associated with non-dietary behaviours, methods of describing diet that incorporate this complexity would be advantageous for use in estimating associations between diet and disease risk. The use of PCA as a tool to capture this complexity has intuitive appeal, but the approach has never been

**Table 1** Foods and/or food groups included in individual discriminant analysis of classification of disease by different food classification schemes

Discriminant analysis	Foods and/or groups associated with each food-use characterisation scheme
(1) 168 individual foods*	Green peppers, red peppers, mushrooms, onions, cucumbers, radishes, celery, lettuce, cauliflower, cooked cabbage, broccoli, Brussels sprouts, sauerkraut, tomatoes, spinach, carrots, broccoli, greens, winter squash, sweet potatoes, asparagus, green beans, yellow beans, corn, peas, beets, lima beans, summer squash, coleslaw, pickles, mashed potatoes, baked potatoes, fried potatoes, other white potatoes, baked beans, tomato juice, cantaloupe, other melon, grapes, berries, apples or apple sauce, pears, peaches, apricots, prunes or plums, pineapple, grapefruit, cherries, oranges or tangerines, bananas, lemons or limes, fruit cocktail, raisins, other dried fruit, olives, apple juice, orange or citrus juice, margarine, butter, mayonnaise or miracle whip, gravy, salad dressing, reduced-calorie mayonnaise, salad dressing, hard cheese, processed cheese, sour cream and dips, half and half, whipped cream, cottage cheese, whole milk, 2% milk, skimmed milk, yoghurt, eggs, cold cereal, oatmeal, other cooked cereal, white bread, white rolls, English muffins or bagels or biscuits, other muffins, pancakes or waffles, French toast, doughnuts or pastries, rice, noodles, bran cereals, dark bread, dark rolls, popcorn, salty snacks, wheat crackers, saltines, ice milk or sherbet, ice cream, cookies, pound or sponge cake, other cake, eclairs or cream puffs, cheese cake, custard or cream pies, fruit pies, pumpkin pies, pudding, brownies, jams or jellies, candy bars, chocolate candy, other candy, peanut butter, peanuts, other nuts, steak, round steak, hamburger patties, other hamburger, other beef, veal, lamb, beef or calves liver, chicken liver, pork roast, pork chops, spareribs, ham, breakfast sausage, sausage, bacon, hot dogs, bologna, liverwurst, other cold cuts, pepperoni, fresh or frozen fish, canned fish, shrimp, other shellfish, chicken wings, fried chicken, other chicken, other poultry, macaroni and cheese, pasta with tomato sauce, pizza, tacos, chilli, pot pies, soup, fast-food cheeseburger, fast-food hamburger, fast-food cheeseburger, fast-food fries, fast-food fried chicken, fast-food fried fish, decaffeinated coffee, regular coffee, hot tea, iced tea, hot cocoa, regular soft drinks, diet soft drinks, beer, wine, liquor
(2) Foods identified from PCA†	
168-item PCA‡	
Healthy	Carrots, pineapple, broccoli, spinach, green peppers, cottage cheese, summer squash, grapefruit, peaches, pears, cauliflower, prunes or plums, mushrooms, beets, grapes, asparagus, berries, other melon, fresh or frozen fish, apples or apple sauce, sweet potatoes, cherries, other white potatoes, raisins
High fat	White rolls, fast-food fries, potato chips, pizza, pasta with tomato sauce, fried potatoes, hot dogs, pepperoni, gravy, hamburger, other beef, doughnuts and pastries, hamburger patties, pork chops, bacon, bologna, other cake, macaroni and cheese, brownies, fast-food cheeseburger, ham, sausage, mashed potatoes, other cold cuts, corn, candy bars
56-item PCA§	
Healthy	Salad, fresh fruit, other vegetables, whole grains, high-carotene vegetables, cruciferous vegetables, canned fruit, high-fat dairy, low-fat dairy, fish and seafood, other white potatoes, baked potatoes, coleslaw, refined grains, desserts
High fat	Processed meats, red meat, pork, fried potatoes, pasta with tomato sauce, pizza, refined grains, snacks, desserts, fast food, macaroni and cheese, mashed potatoes, candy, low-calorie salad dressing
36-item PCA¶	
Healthy	Vegetables, fruit, whole grains, low-fat dairy, refined grains, high-fat dairy, desserts, fish or seafood, nuts, low-calorie fats, jams and jelly, eggs
High fat	Processed meats, meat, pizza, pasta with tomato sauce, snacks, refined grains, fast foods, desserts, macaroni and cheese, candy, tacos, low-calorie fats
(3) Fruits and vegetables†	Green and red peppers, mushrooms, onions, cucumbers, radishes, celery, lettuce, cauliflower, cabbage, broccoli, Brussels sprouts, sauerkraut, tomatoes, spinach, carrots, broccoli, greens, winter squash, sweet potatoes, asparagus, green and yellow beans, corn, peas, beets, lima beans, summer squash, coleslaw, mashed potatoes, baked potatoes, fried potatoes, other white potatoes, baked beans, tomato juice, cantaloupe and other melon, grapes, berries, apples, pears, peaches, apricots, prunes or plums, pineapple, grapefruit, cherries, oranges or tangerines, bananas, lemons or limes, apple sauce, fruit cocktail, raisins, other dried fruit, apple juice, orange or citrus juice
(4) USDA food groups*	Fresh fruit, canned or frozen fruit, raw vegetables, cooked vegetables, red meat, poultry, fish and seafood, processed meats, snacks and sweets, grain products, dairy, and fats

\* Stepwise selection procedure.

† Items forced as a block into discriminant analysis.

‡ PCA of 168 individual foods and beverages.

§ PCA of 56 detailed food groups, foods and beverages.

¶ PCA of 36 broad food groups and individual foods and beverages.

compared to less statistically sophisticated methods of characterising diet. To be successful at characterising important differences in dietary behaviour associated with risk of disease, the method used should maximally

discriminate subjects by disease status, thereby providing a clearer delineation of the foods most strongly related to risk. Our results suggest that the use of PCA may not provide substantial advantages over more commonly

**Table 2** Percentage of subjects correctly classified as cases and controls by specific patterns of food use

	Cases	Controls
Stepwise discriminant analysis*	64.7	63.1
'Healthy' pattern from PCA		
168 items†	65.5	60.7
56 items†	55.6	55.2
36 items†	54.7	54.0
'High fat' pattern from PCA		
168 items†	66.8	57.1
56 items†	60.3	54.9
36 items†	59.9	49.5
Fruits and vegetables	67.7	62.9
Stepwise food groups‡	57.3	59.2

\* Stepwise discriminant analysis identified from the original 168 items the following 11 foods as maximally separating cases and controls: saltines, corn, spinach, butter, olives, soup, bananas, tacos, beets, pancakes, and coffee.

† The number of items entered into PCA. Only those foods identified by PCA for each pattern and method were included in the specific discriminant analyses.

‡ Stepwise discriminant analysis identified from 12 food groups (USDA classification) the following three groups as maximally separating cases and controls: non-raw fruits, non-raw vegetables, and grains.

used less sophisticated methods of characterising diet. Furthermore, for the PCA-derived dietary patterns, greater detail in characterising food use produced a higher percentage of subjects correctly classified on disease status.

Principal components analysis is a sophisticated statistical technique involving a fair amount of data manipulation. Although these methods have been applied in studies of diet and disease, no studies to date have assessed the improvement in characterisation of diet by this method compared with use of food groups or single foods. Our study suggests that simpler methods of characterising diet can discriminate subjects on disease status similarly to PCA-derived dietary patterns. In fact, fruits and vegetables were identified as foods that provided the highest correct classification of cases and controls. This finding is not surprising given that intake of any food or food group is dependent on intake of other foods or groups. Total intake tends to be tightly controlled; individuals who eat greater quantities of fruit and vegetables also tend to eat less of other food groups<sup>15</sup>. Finally, the items in the fruit and vegetable group consisted primarily of the foods identified in the 'healthy' patterns, providing further evidence that simpler methods may be sufficient in describing dietary behaviour.

Of interest is the finding that none of the methods of characterising diet produced correct classifications that could be considered remarkable, although correct classifications between 60% and 70% are generally considered good<sup>14</sup>. Furthermore, the differences in correct classification between the four food-use characterisations were

fairly small, suggesting that the ability to differentiate between cases and controls was similar for the four methods examined. Although the use of stepwise discriminant analysis to obtain an objective set of foods most predictive of case-control status produced correct classification similar to the other diet characterisations, the foods identified failed to represent dietary patterns interpretable within our present conceptual framework. Whereas our results indicate few differences in the ability of different diet characterisation methods to describe the diets of diseased and non-diseased subjects, they also suggest that challenges remain in the field of dietary assessment methodology.

The use of endometrial cancer as an outcome might have affected the differences in correct classification obtained from the different methods of characterising dietary intake. However, the intent of the present analyses was not to estimate risk, but to compare different methods of describing diet as to their ability to correctly identify subjects as diseased or not. It is unlikely that the weak association between endometrial cancer and diet had an impact on these analyses as all of the comparisons were internal and the different food characterisations were obtained using the same set of food-frequency data. Although dietary factors are not strongly related to endometrial cancer risk, the literature has been fairly consistent in showing reduced risk of this cancer associated with higher fruit and vegetable intake<sup>15</sup>. In fact, the highest correct classification of cases and controls was obtained for the 47 fruits and vegetables, providing further evidence that these foods may indeed be important in the aetiology of endometrial cancer. Finally, to address the concern that a weak diet-disease relationship might bias our results, we repeated the present analyses using dietary data from a large, population-based case-control study of myocardial infarction and obtained results similar to those reported here (data not shown).

In conclusion, our study suggests that PCA-derived dietary patterns provide information about dietary behaviour similar to simpler methods of characterising diet, but do not offer a substantial improvement in discrimination of individuals on disease status. In fact, we found that fruits and vegetables alone provided the highest (albeit, not dramatic) discrimination among endometrial cancer cases and controls compared with the other methods of characterisation. Sophisticated techniques may be unnecessary in studies of diet and disease when a small number of foods adequately describes dietary behaviour in association with disease.

## References

- 1 Randall E, Marshall JR, Graham S, Brasure J. Frequency of food use data and the multidimensionality of diet. *J. Am. Diet Assoc.* 1989; **89**: 1070-5.
- 2 Randall E, Marshall JR, Graham S, Brasure J. High risk health

- behaviors associated with various dietary patterns. *Nutr. Cancer* 1991; **16**: 135–51.
- 3 Ursin G, Ziegler RG, Subar AF, Graubard BI, Haile RW, Hoover R. Dietary patterns associated with a low-fat diet in the national health examination follow-up study: identification of potential confounders for epidemiologic analyses. *Am. J. Epidemiol.* 1993; **137**(8): 916–27.
  - 4 Randall E, Marshall JR, Graham S, Brasure J. Patterns in food use and their associations with nutrient intakes. *Am. J. Clin. Nutr.* 1990; **52**: 739–45.
  - 5 Randall E, Marshall JR, Brasure J, Graham S. Patterns in food use and compliance with NCI dietary guidelines. *Nutr. Cancer* 1991; **15**: 141–56.
  - 6 Randall E, Marshall JR, Brasure J, Graham S. Dietary patterns and colon cancer in western New York. *Nutr. Cancer* 1992; **18**: 265–76.
  - 7 Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN. Eating patterns and risk of colon cancer. *Am. J. Epidemiol.* 1998; **148**(1): 4–16.
  - 8 Hu FB, Rimm EB, Stampfer MJ, Ascherio A, Spiegelman D, Willett WC. A prospective study of major dietary patterns and risk of coronary heart disease in men. *Am. J. Clin. Nutr.* 2001; **72**(4): 912–21.
  - 9 Hu FB, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, Sampson L, Willett W. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *Am. J. Clin. Nutr.* 1999; **69**: 243–9.
  - 10 Willett WC. *Nutritional Epidemiology*, 2nd edn. New York: Oxford University Press, 1998.
  - 11 Johnson DE. *Applied Multivariate Methods for Data Analysts*. Pacific Grove, CA: Duxbury Press, 1998.
  - 12 McCann SE, Freudenheim JL, Marshall JR, Vena JE, Laughlin R, Brasure JR, Swanson MK, Graham S. Diet in the epidemiology of endometrial cancer in Western New York. *Cancer Causes Control* 2000; **11**(10): 965–74.
  - 13 Logue AW, Smith ME. Predictors of food preferences in adult humans. *Appetite* 1986; **7**: 109–25.
  - 14 Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. Norwalk, CT: Appleton & Lange, 1993.
  - 15 American Institute for Cancer Research. Endometrium. In: *Food, Nutrition and the Prevention of Cancer: A Global Perspective. Part II: Cancers, Nutrition and Food*, Chapter 4.13. Washington, DC: American Institute for Cancer Research, 1997.