Nutritional Sciences 7(1), February, 2004, pp. 53-58

The Multiethnic Cohort Study of Diet and Cancer: Design and Early Findings

Laurence N. Kolonel[§]
Cancer Research Center, University of Hawaii, USA

The Multiethnic Cohort Study was designed to study prospectively the relationship of diet and other lifestyle factors to the risk of cancer. The cohort was established in 1993-1996 in Hawaii and California (primarily Los Angeles) and included a representative sample of more than 215,000 men and women primarily from five different ethnic groups: Japanese, whites, Native Hawaiians, Latinos, and African- Americans. Because of the emphasis on diet, great attention was paid to developing and pre-testing a self-administered quantitative food frequency questionnaire that would adequately assess food and nutrient intakes in these groups. An extensive food composition database was also created for the study. In addition, during data collection, a calibration study was conducted that makes possible adjustment for measurement error in nutrient intakes and valid comparison of intakes across the several ethnic groups. At the present time, blood and urine specimens are being collected from cohort participants and should yield a biorepository of more than 80,000 subjects. Baseline data indicate that the cohort is well representative of the general population of Hawaii and California, so that results can be generalized. These data also show a wide range in dietary intakes and in other lifestyle variables that should facilitate the testing of etiologic hypotheses.

Key words : Multiethnic cohort study, Diet, Lifestyle, Cancer

INTRODUCTION

The ability to assess dietary intake in large population samples is essential for conducting prospective epidemiologic studies of nutrition and cancer. The problem is especially challenging when the investigation involves several different population groups with varying eating patterns. This adds complexity to the development of the assessment tool (e.g., a self-administered dietary questionnaire), as well as to the creation of a food composition database for analysis of the data. The Multiethnic Cohort Study (MEC) is an example of a project that depends heavily on such nutritional assessment and the use of food and nutrient databases for epidemiologic research. This study was established in Hawaii and California (primarily Los Angeles) between 1993 and 1996 in order to investigate the relationship between diet and cancer risk. Many years of research had preceded the setting up of this study, in which the advantages of the diverse ethnic populations in Hawaii (and more recently, California) for investigating etiologic hypotheses had been demonstrated (e.g.,1-5). These studies suggested that a prospective multiethnic cohort which would examine dietary and non-dietary variables measured by questionnaire or by analysis of biologic specimens could substantially advance our understanding of the causes of cancer and the means to prevention.

STUDY DESIGN

The MEC was designed to provide a wide range of dietary intakes by including several different ethnic groups with varying eating patterns[6]. Based on a number of considerations, including the size of the different ethnic groups in Hawaii and Los Angeles, anticipated response rates, and known risks of the major cancers of interest, adults (age 45-75) from the following five groups were selected to comprise the cohort: Japanese Americans, whites, Native Hawaiians (Polynesians), African Americans and Latinos. The study was also designed to take advantage of the population-based cancer registries in Hawaii and California which, by computer linkage of the cohort and registry files, would facilitate surveillance for cancer incidence. In order to have as representative a sample as possible, Drivers License Files were used as the primary source of potential participants in the study, since most adults in the U.S. are drivers. Nevertheless, this source was supplemented with some additional files for even better representation.

Accepted : January 26, 2004
§ To whom correspondence should be adressed.

QUESTIONNAIRE DEVELOPMENT

The questionnaire for the MEC emphasized the assessment of food intake and was designed specifically for this project. Initially, 3-day measured food records were collected from approximately 60 men and women (45 to 75 years of age) of each ethnic group in order to identify food items for inclusion in the questionnaire. For each ethnic group, the contribution of each food item to the total intake of the nutrients of major interest (e.g., fat, dietary fiber, vitamin A, carotenoids, and vitamin C), and the minimum set that yielded at least 85% of the intake for a particular nutrient was determined. Because this was done for each ethnic group and nutrient, the final questionnaire list actually accounts for much more than 85% of the intake of the major nutrients. In addition, the food records were used to identify for inclusion specific food items uniquely associated with the traditional diets of a particular group irrespective of their contribution to intake (e.g., tofu and salted fish for Japanese; tortillas and tamales for Latinos). Food items specific to the diets of other ethnic groups in Hawaii and Los Angeles, such as Filipinos, Chinese and Koreans, were not included in the questionnaire, unless they were commonly eaten by other ethnic groups (e.g., kim chee, chow mein noodles).

The questionnaire was designed for self-administration and for processing by optical scanning. Eight frequency categories for foods and nine for beverages were included, together with three (in a few instances four) choices of portion size. As an additional aid to quantification, photographs showing selected foods in representative portion sizes were provided at the head of several pages of the questionnaire. The collection of portion size information was important because of significant variations among individuals as well as significant variations in the typical serving sizes consumed by different ethnic groups. The portion size options on the questionnaire were based on typical serving sizes for each single food or grouping of foods as reflected in the original 3-day measured food records.

In addition to the section on diet, the questionnaire included information on demographic factors (including migrant status), personal behaviors (smoking, solar exposure, physical activity), history of prior medical conditions, use of medications, reproductive history and the use of replacement estrogens/oral contraceptives (for women), and a brief family history of cancer.

The questionnaire was extensively pretested, culminating in the mailing of a draft questionnaire to 1,000 randomly selected subjects of each ethnic group. In addition, the reproducibility of the questionnaire was tested on approximately 300 subjects to whom the form was mailed again six months later; correlation coefficients for nutrient intakes between the two questionnaires were in the range of 0.5-0.7 for most items. Finally, unannounced 24-hour recalls were collected monthly over a 12-month interval from a sample of 337 subjects representing all ethnic-sex groups who had completed the pilot questionnaire; these data were used to design a calibration study for the final dietary protocol (see below).

DATA COLLECTION, EDITING AND MANAGEMENT

The cohort was assembled between 1993 and 1996. The self-administered, 26-page questionnaire was mailed (up to three times) to persons between the ages of 45 and 75 in the entire state of Hawaii and to several counties in California. Response rates varied by ethnicity as follows: Japanese men 46.3%, women 51.3%; white men 39.1%, women 47.0%; African American men 20.0%, women 25.5%; Latino men 18.6%, women 21.3%; and Native Hawaiian men 36.1%, women 42.4%.

After some initial manual editing, the questionnaires were processed by optical scanning and the information was stored in a structured database. This was followed by computer-assisted editing for such problems as multiple responses to a single item, inconsistencies between parts of the questionnaire and other correctable errors. The study is coordinated from the Hawaii center, where the computer management and nutritional activities take place. A computerized tracking system and food composition database specific for this study were developed and are being maintained by the coordinating center. The complex tracking database allows for recording of all addresses, as well as tumor, death and other follow-up information. However, the staff at each location maintain current addresses and conduct follow-up activities for the cohort members in their area.

FOOD COMPOSITION TABLE

An extensive food composition database for use in the multiethnic population of Hawaii has been developed and maintained at the Cancer Research Center of Hawaii for many years. For the present project, additional data on foods consumed by Latinos and African Americans had to be added. These data were obtained from USDA Handbook 8[7,8], with supplementation from special laboratory analyses (unpublished data), and other research and commercial publications[9-12]. This database includes over 1,500 foods and 700 recipes, the latter developed from food records and 24-hour recalls coll-

ected over many years, as well as from established cookbooks. This extensive data base is constantly reviewed and updated.

For the analysis of the self-administered dietary questionnaire for this study, a customized food composition table, based on the larger data base, was prepared. For grouped (composite) food items on the questionnaire, the individual foods were weighted according to their respective frequencies of consumption as determined from the 24-hour recalls collected on the first 1,362 participants in a calibration study.

CALIBRATION STUDY

Because recall information, especially in the assessment of dietary intakes, can entail considerable measurement error[13], we collected three unannounced 24-hour diet recalls about one month apart on approximately 260 randomly-selected subjects of each ethnic-sex group in the study. Day of the week was randomly selected for the second and third recalls to ensure an adequate representation across days of the week. These data were used to develop calibration equations that could reduce the misclassification error in the estimation of risk associated with nutrient intakes[14,15]. By applying these equations, valid inter-ethnic comparisons could be made even if the questionnaire worked better in some ethnic groups than in others. The design and findings of the calibration study has been reported[16].

COLLECTION OF BIOLOGIC SPECIMENS

For logistic and financial considerations, biologic specimens were not collected at the time the baseline questionnaire was completed by the participants. However, funds for setting up a prospective biorepository were obtained subsequently, and a substantial effort began in 2002 to collect blood and urine samples from more than 80,000 members of the cohort. Blood samples are being stored in 0.5cc aliquots under liquid nitrogen. Urine samples are being stored in 2.0cc aliquots at -80℃.

BASELINE FINDINGS IN THE COHORT

Table 1(a,b) shows the distribution of the cohort by age, sex and ethnicity. There are substantial numbers of participants from all of the five ethnic groups selected for the study, though the number of Native Hawaiians is necessarily lower because of the smaller size of this population group. All age groups in the range selected for the study are also well represented. Table 2 shows that the cohort includes many first generation migrants, especially in the Latino group, thus affording the distinct opportunity to conduct migrant studies within the cohort. Table 3 compares the distribution of male cohort members on educational level with the general population of Hawaii and California (based on data from the 1990 U.S. census). These results attest to the representativeness of the sample included in the study. Similar results were seen for females.

Table 1. Distribution of the Cohort by Age*, Sex, and Ethnicity

a. Males

Ethnicity	Age Group			
	45-54	55-64	65-75	Total No.
Latino	6,376	9,453	6,989	22,818
African-American	2,899	3,503	6,449	12,851
Japanese	8,065	7,667	11,232	26,964
White	8,718	6,672	7,467	22,857
Native Hawaiian	2,779	1,969	1,372	6,120
Other	1,744	1,801	1,655	5,200
Total	30,581	31,065	35,164	96,810

* Age at baseline (completion of the QFFQ)

b. Females

Ethnicity	Age Group			
	45-54	55-64	65-75	Total No.
Latino	7,767	10,481	6,372	24,620
African-American	6,136	6,315	9,805	22,256
Japanese	8,809	9,299	11,849	29,957
White	10,216	7,937	8,349	26,502
Native Hawaiian	3,753	2,426	1,672	7,851
Other	2,949	2,495	1,811	7,255
Total	39,630	38,953	39,858	118,441

* Age at baseline (completion of the QFFQ)

Table 2. Percent Distribution of Cohort Members by Migration Status

Ethnicity	Foreign-born	U.S. - born	
		2nd generation*	Later generation
Latino	52.0**	22.3	25.7
African-American	3.1	0.7	96.2
Japanese	8.4	44.0	47.6
White	11.6***	7.6	80.8
Native Hawaiian	0.3	0.4	99.2
Other	61.6	18.2	20.2
Total	20.4	19.5	60.0

* Both parents foreign-born
** 68% of these were born in Mexico
*** 47% of these were born in Europe

Table 4 shows the distribution of the cohort on certain lifestyle and diet-associated variables. Smoking rates vary by ethnicity, with the highest rates in African

Table 3. Distribution (%) of the Cohort Members by Educational Level and Comparison with the U.S. Census[1] (males)

Ethnicity	≤ 8 years		9-12 years		Vocational		At least some college	
	Cohort	Census	Cohort	Census	Cohort	Census	Cohort	Census
Latino	33.6	43.2	31.1	36.8	6.6	2.0	28.7	18.0
African- American	9.3	13.0	33.7	47.5	5.5	3.1	51.5	36.4
Japanese	3.2	7.9	34.7	40.6	12.6	7.2	49.4	44.3
White	3.4	2.9	19.7	30.8	4.2	3.8	72.7	62.5
Native Hawaiian	5.8	7.0	45.9	61.4	8.5	4.3	39.9	27.4

1) Census data for the corresponding geographic areas of Hawaii and California, 1990
Source: Bureau of the Census. 1990 Census Population--General Population Characteristics

Table 4. Smoking, Drinking, Obesity, and Physical Activity in the Cohort by Sex and Ethnicity (%)[1]

Ethnicity	Current Smokers	Current Drinkers[2]	BMI > 30.0	Vigorous Physical Activity[3]
MALES				
Latino	18.3	40.5	21.6	36.1
African American	28.8	37.9	22.8	27.5
Japanese	13.3	38.6	7.3	28.2
White	19.8	56.0	14.6	40.1
Native Hawaiian	21.8	38.5	35.1	47.1
FEMALES				
Latino	13.6	13.7	29.5	12.0
African American	20.9	17.6	37.7	9.3
Japanese	11.1	9.9	6.7	9.6
White	17.6	35.3	19.6	19.5
Native Hawaiian	21.2	16.2	34.5	23.1

1) Standardized to the overall cohort age distribution
2) ≥ 1 beverage/week
3) ≥ 3 hrs of strenuous exercise or vigorous work/week

American men and in Native Hawaiian women. White men and women have the highest proportion of alcohol consumers. The proportion of subjects who are obese (i.e., BMI greater than 30) is extremely high for the Latino, African American and Native Hawaiian groups; only the Japanese show low rates of obesity. The Native Hawaiian group had the highest proportion of participants who engage in regular vigorous physical activity, whereas the African Americans had the lowest.

Some dietary findings are shown in the next tables. Table 5 shows correlations for men between the 24-hour recalls and the food frequency questionnaire for selected nutrient densities based on the calibration study; the findings for women are very similar (not shown). The correlations are quite comparable among the ethnic groups within sex, though the questionnaire performed somewhat better in whites than in the other groups. Table 6 shows the intakes of selected foods by ethnicity and sex. Some notable findings include the much higher consumption of legumes by the Latinos, the much lower consumption of dairy products by the Japanese, and the higher consumption of fish by the Native Hawaiians. Selected nutrient intakes are shown in Table 7 and reflect these different eating patterns. Fiber and calcium intake

Table 5. Correlations between 24-Hour Recalls and Food Frequency Questionnaire (MALES)

Nutrient	African Americans	Japanese Americans	Latinos	Whites
Fat, total[1]	0.60	0.68	0.70	0.69
Saturated	0.71	0.71	0.68	0.72
Carbohydrate[1]	0.47	0.54	0.57	0.68
Protein[1]	0.44	0.46	0.35	0.43
Ca^{++} (mg)[2]	0.65	0.72	0.46	0.65
Vit C (mg)[2]	0.64	0.74	0.65	0.72
β-carotene (mcg)[2]	0.67	0.63	0.68	0.70
Fiber (g)[2]	0.69	0.79	0.68	0.71

1) % calories
2) per 1,000 Kcal

are highest among Latinos, and lowest among Japanese.

Table 8 illustrates how the cohort can be used to address ethnic differences in cancer risk. As seen in first line of the table, African Americans and Latinos (first as well as second generation) have lower rates of breast cancer compared with white women. However, after additional adjustment for the major known risk factors for this cancer, the risk in African Americans and Latinos becomes similar to that in whites, indicating that the ethnic discrepancies in incidence can be largely accounted

Table 6. Intake of Selected Foods[1] from the QFFQ among the Main Ethnic Groups in the Cohort

Ethnicity	Red Meat	Fish	Dairy Products	Legumes (excl. Tofu)	Cruciferous Vegetables
MALES					
Latino	72.9	17.2	277.8	96.4	34.2
African American	54.4	20.6	202.6	38.9	35.6
Japanese	49.0	32.9	135.2	24.2	44.9
White	51.2	25.1	257.9	34.4	40.4
Native Hawaiian	68.4	39.4	214.6	27.4	46.4
FEMALES					
Latino	54.0	14.0	273.1	67.4	38.3
African American	38.5	18.6	201.8	31.6	47.0
Japanese	36.9	23.5	153.8	20.4	49.2
White	35.0	17.6	265.7	26.5	43.2
Native Hawaiian	52.6	32.0	237.1	24.5	51.4

1) Mean daily intake (g), standardized to the overall cohort age distribution.

Table 7. Energy and Nutrient Intakes[1] from the QFFQ among the Main Ethnic Groups in the Cohort

Ethnicity	Calories(kcal)	Total Fat(g)	Dietary Fiber(g)	Vitamin C(mg)	Calcium(mg)
MALES					
Latino	2679.0	98.1	33.5	194.9	1056.2
African6 American	2278.1	85.7	25.5	167.8	771.0
Japanese	2241.8	70.1	21.0	152.5	634.6
White	2340.4	80.9	26.4	182.9	882.1
Native Hawaiian	2727.6	91.6	25.9	192.0	834.8
FEMALES					
Latino	2285.5	82.8	30.6	211.7	965.3
African American	1966.1	72.5	25.6	187.3	745.8
Japanese	1864.9	59.6	22.2	180.5	633.8
White	1899.4	65.3	24.5	180.0	843.8
Native Hawaiian	2348.7	79.6	27.4	218.4	847.4

1) Mean daily intake, standardized to the overall cohort age distribution.

or by known variables. However, for Native Hawaiians, the higher initial risk becomes greater with the additional adjustment. Thus, other factors must be involved in the very high incidence of breast cancer in this ethnic group[17].

Table 8. Comparative Breast Cancer Risk Adjusted for Known Risk Factors

Relative Risk	White	African American	Native Haw'n	Japanese	Latino US-Born	Latino Non-US
Adj.(1)[1]	1.00	0.78	1.33	0.99	0.77	0.60
Adj.(2)[2]	1.00	0.98	1.65	1.11	0.95	0.84

1) Adjusted for age
2) Adjusted for age and 7 additional breast cancer risk factors (age at menarche, age at first birth, parity, age at and type of menopause, weight, HRT use, and alcohol consumption).

Table 9 shows an example of the use of the biological specimens in the cohort to examine genetic susceptibility. The figure shows that men who inherit a missense substitution in the 5alpha-reductase Type II gene (SRD5A2) are at higher risk for prostate cancer, especially advanced disease[18]. Considerable work is currently underway in the MEC along such lines, with recent efforts examining haplotypes.

Table 9. The A49T Missense Substitution in 5α-Reductase Type II *(SRD5A2)* and Prostate Cancer Risk in African-American and Latino Men

	OR for AT or TT Genotype		
	Controls	Cases	
		Localized	Advanced
African-Americans	1.0	1.5	7.2*
Latinos	1.0	1.7	3.6*

* P < .05

SUMMARY

The Multiethnic Cohort Study offers several advantages for research on the etiology of cancer. 1) The wide range in exposures facilitates the investigation of dietary and other lifestyle relationships to cancer. 2) The diverse composition of the cohort holds considerable potential to yield insights into reasons for ethnic variations in

cancer risk. 3) The findings can be generalized, because the cohort is quite well representative of the general population. 4) The cohort design permits researchers to investigate many disease endpoints besides cancer. 5) The collection of biospecimens will enable investigators to study genetic markers of risk and, ultimately, gene-environment interactions. 6) The stored biospecimens will also permit researchers to study biomarkers of nutritional, hormonal and other exposures in relation to various disease outcomes.

Literature Cited

1) Hirohata T, Nomura AM, Hankin JH, Kolonel LN, Lee J. An epidemiologic study on the association between diet and breast cancer. *J Natl Cancer Inst.* 78:595-600, 1987
2) Hinds MW, Stemmermann GN, Yang HY, Kolonel LN, Lee J, Wegner E. Differences in lung cancer risk from smoking among Japanese, Chinese and Hawaiian women in Hawaii. *Int J Cancer* 27:297-302, 1981
3) Kolonel LN, Hankin JH, Lee J, Chu SY, Nomura AMY, Hinds MW. Nutrient intakes in relation to cancer incidence in Hawaii. *Br J Cancer* 44:332-9, 1981
4) Le Marchand L, Yoshizawa CN, Kolonel LN, Hankin JH, Goodman MT. Vegetable consumption and lung cancer risk: a population-based case-control study in Hawaii. *J Natl Cancer Inst* 81:1158-64, 1989
5) Le Marchand L, Wilkens LR, Hankin JH, Kolonel LN, Lyu LC. A case-control study of diet and colorectal cancer in a multiethnic population in Hawaii (United States): lipids and foods of animal origin. *Cancer Causes Control* 8:637-48, 1997
6) Kolonel LN, Henderson BE, Hankin JH, Nomura AMY, Wilkens LR, Pike MC, Stram DO, Monroe KR, Earle ME, Nagamine FS. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 151:346-57, 2000
7) US Department of Agriculture. USDA nutrient database for standard reference. Release no. 10. Riverdale, MD: Nutrient Data Laboratory, Agricultural Research Service, 1993
8) US Department of Agriculture. USDA nutrient database for standard reference. Release no. 7. Riverdale, MD: Nutrient Data Laboratory, Agricultural Research Service, 1991
9) Pennington JAT. Bowes & Church food values of portions commonly used. 16th ed. Philadelphia, PA: Lippincott, 1994.
10) Kagawa R. Standard tables of food composition in Japan. Tokyo, Japan: Women's University, 1995.
11) Chavez MM, Hernandez M, Roldan JA. Tablas de uso practico del: valor nutritivo de los alimentos de mayor consume en Mexico. Mexico: Instituto Nacional de la Nutricion Salvador Zabiran, 1992
12) Holland B, Welch AA, Unwin ID, et al. McCance and Widdowson's The Composition of Foods. 5th ed. Cambridge, UK: Royal Society of Chemistry and Ministry of Agriculture, Fisheries and Food, 1991
13) Willett W. Nutritional epidemiology. New York: Oxford University Press, 1990
14) Stram D, Longnecker M, Shames L, et al. Cost efficient design of a diet validation study. *Am J Epidemiol* 142: 353-6, 1995
15) Kaaks R, Riboli E, van Staveren W. Calibration of dietary intake measurements in prospective cohort studies. *Am J Epidemiol* 142:548-56, 1995
16) Stram DO, Hankin JH, Wilkens LR, Pike MC, Monroe KR, Park S, Henderson BE, Nomura AMY, Earle ME, Nagamine FS, Kolonel LN. Calibration of the dietary questionnaire for a multiethnic cohort in Hawaii and Los Angeles. *Am J Epidemiol* 151:358-70, 2000
17) Pike MC, Kolonel LN, Henderson BE, Wilkens LR, Hankin JH, Feigelson HS, Wan PC, Stram DO, Nomura MY. Breast cancer in a multiethnic cohort in Hawaii and Los Angeles: risk factor-adjusted incidence in Japanese equals and in Hawaiians exceeds that in whites, *Cancer Epidemiol Biomarkers Prev* 11:795-800, 2002
18) Makridakis N, Ross RK, Pike MC, Chang L, Stanczyk FZ, Kolonel LN, Shi CY, Yu MC, Henderson BE, Reichardt JK. A prevalent missense substitution that modulates activity of prostatic steroid 5alpha-reductase. *Cancer Res* 57:1020-1022, 1997