The American Cancer Society Cancer Prevention Study II Nutrition Cohort

Rationale, Study Design, and Baseline Characteristics

Eugenia E. Calle, Ph.D.
Carmen Rodriguez, M.D.
Eric J. Jacobs, Ph.D.
M. Lyn Almon, M.S.P.H.
Ann Chao, Ph.D.
Marjorie L. McCullough, Sc.D.
Heather S. Feigelson, Ph.D.
Michael J. Thun, M.D.

American Cancer Society, Atlanta, Georgia.

This article was previously published (*Cancer* 2002;94:500–511); the earlier published version did not reflect key edits which were necessary in light of the authors' decision about when to terminate the follow-up of the cohort. Readers are asked to consider this reprinted version of the article to be the official citation.

The authors acknowledge the hundreds of thousands of ACS volunteers whose trust and goodwill make epidemiologic research at the ACS possible. They also thank the many ACS staff members who have gone before, including Dr. E. Cuyler Hammond, Lawrence Garfinkel, and Dr. Clark W. Heath, Jr., upon whose foundation they continue to build. Finally, the authors thank the Illinois Division for generous support of the CPS-II Nutrition and Life-Link Cohorts.

Address for reprints: Eugenia E. Calle, Ph.D., American Cancer Society, 1599 Clifton Road, N.E., Atlanta, GA 30329; Fax: (404) 327-6450; E-mail: jcalle@cancer.org

Received June 7, 2001; revision received September 7, 2001; accepted September 11, 2001.

BACKGROUND. Large-scale, prospective cohort studies have played a critical role in discovering factors that contribute to variability in cancer risk in human populations. Epidemiologists and volunteers at the American Cancer Society (ACS) were among the first to establish such cohorts, beginning in the early 1950s and continuing through the present, and these ACS cohorts have made landmark contributions in many areas of epidemiologic research.

METHODS AND RESULTS. The Cancer Prevention Study II Nutrition Cohort was established in 1992 and was designed to investigate the relation between diet and other lifestyle factors and exposures and the risk of cancer, mortality, and survival. The cohort includes over 84,000 men and 97,000 women who completed a mailed questionnaire in 1992. New questionnaires are sent to surviving cohort members every other year to update exposure information and to ascertain new occurrences of cancer; a 90% response rate was achieved for follow-up questionnaires in 1997 and 1999. Reported cancers are verified through medical records, registry linkage, or death certificates. The cohort is followed actively for all cases of incident cancer and for all causes of death. Through a collaborative effort among ACS national and division staff, volunteers, and the American College of Surgeons, blood samples were collected from a subgroup of 40,000 cohort members and are in storage at a central repository for future investigation of dietary, hormonal, genetic, and other factors and cancer risk. Collection of DNA samples from buccal cells in an additional 50,000 cohort members is underway currently and will be completed in 2002.

CONCLUSIONS. This new cohort of both men and women promises to be particularly valuable for the study of cancer occurrence, mortality, and survival as they relate to obesity and weight change, physical activity at various points in life, vitamin supplement use, exogenous hormone use, other medications (such as aspirin and nonsteroidal anti-inflammatory drugs) and cancer screening modalities. ***Cancer* 2002;94:2490–2501.**
DOI 10.1002/cncr.101970

KEYWORDS: biomarkers, cohort studies, diet, hormones, obesity, physical fitness, prospective studies, screening, smoking.

The evolution of the modern cohort study has received particular attention recently.[1,2] Early cohorts typically addressed very specific exposure issues and generally were limited in their ability to control for confounding. More recently, large cohort studies have collected questionnaire information and biospecimens to assess a comprehensive array of lifestyle, medical, and other risk factors as potential determinants of a wide range of specific diseases.

Newer cohorts have taken advantage of structured dietary questionnaires developed in the 1980s and early 1990s that can be admin-

TABLE 1
Characteristics of American Cancer Society Cohorts

Characteristic	Hammond-Horn	CPS-I	CPS-II		
			Baseline	Nutrition	LifeLink
Study design					
Study participants	187,783	1,051,038	1,185,106	184,194	39,200
Volunteer recruiters	22,000	68,000	77,000	7000	2000
Yrs of follow-up	1952–1955	1959–1972	1982 (ongoing)	1992 (ongoing)	1998 (ongoing)
No. of states	9	25	50	21	20
Outcome of interest	Mortality	Mortality	Mortality	Incidence and mortality	Incidence and mortality
Demographics					
Women (%)	0	57	57	53	56
White (%)	100	97	93	97	98
Black (%)	0	2	4	1	1
Married (%)	?[a]	85	83	88	94
College graduate (%)	?[a]	17	30	38	46
Age at entry (median) (yrs)	?[a]	52	57	63	70
Age range (yrs)	50–69	30–108	30–111	40–92	49–96
Current cigarette smokers (%)	57	29	20	7	3

CPS-II: Cancer Prevention Study II.
[a] Data from the Hammond–Horn cohort have been lost.

istered to large groups of people and that discriminate among people with respect to dietary intake.[3] More than 30 prospective cohorts with a focus on dietary assessment have been assembled to date; taken together, these cohorts include over 3 million people, primarily from North America and Europe.[3]

It is interesting to note that, in contrast to early cohorts, women substantially outnumber men in most of these contemporary cohorts. Only four cohorts have enrolled 80,000 or more men; these include the EPIC multicenter cohort study conducted in nine European countries,[4] the Multiethnic Cohort study conducted in African-American, Latino, Japanese, Native Hawaiian, and white populations living in Hawaii and California,[5] the National Cancer Institute Diet and Health Study conducted in members of the American Association of Retired Persons,[3] and the American Cancer Society (ACS) Nutrition Cohort conducted in middle-aged to elderly men and women living in 21 states in the United States.[6] In this article, we present the rationale, study design, data collection, and descriptive characteristics of the ACS Nutrition Cohort.

MATERIALS AND METHODS

History

In 1952, the ACS began its first large-scale prospective cohort study, the Hammond–Horn Study,[7,8] to examine the effect of cigarette smoking on death rates from cancer and other diseases (Table 1). This landmark research effort was one of the first cohort studies to establish the association of smoking with lung carcinoma, heart disease, and all-cause mortality; it set the methodologic foundation for the two subsequent Cancer Prevention Study cohorts (CPS-I and CPS-II).

Compared with the Hammond–Horn Study, both CPS-I and CPS-II were designed to address a wide range of potential exposures, in addition to tobacco use, that may be associated with cancer (Table 1). CPS-I data have resulted in the publication of more than 100 scientific articles, including analyses of smoking,[9–12] body weight,[13–18] air pollution,[19] and risk factors for breast carcinoma[20–22] in relation to mortality. Whereas the Hammond–Horn data have been lost, CPS-I data continue to be analyzed.

CPS-II was launched in 1982 using volunteers to enroll approximately 1.2 million men and women in all 50 states, the District of Columbia, and Puerto Rico (Table 1).[23] Participants completed a four-page confidential questionnaire that included questions on demographic characteristics, height and weight, personal and family history of cancer and other diseases, use of medicines and vitamins, reproductive history and hormone use (women), occupational exposures, brief dietary and physical activity assessments, and alcohol and tobacco use. Exposure information for the entire CPS-II cohort was collected only once at baseline in 1982. Mortality follow-up is ongoing and is accomplished through linkage with the National Death Index[24]; 283,600 deaths have occurred in the cohort through December 31, 1998. Over 100 publications from CPS-II address a broad range of lifestyle factors, including alcohol use,[25] exogenous hor-

mones,[26–30] aspirin,[31,32] vitamin supplements,[33,34] air pollution,[35] tobacco use,[36–43] family history of cancer,[44–46] occupation,[47,48] and adiposity.[49–52] Analysis of CPS-II data in relation to mortality is ongoing. Complete citations for all publications can be viewed on the ACS website at www.cancer.org.

CPS-II Nutrition Cohort

Selection of participants

The CPS-II Nutrition Cohort was established in 1992 and 1993 as a subgroup of the larger CPS-II Cohort with three primary objectives: 1) to obtain detailed information on dietary and other exposures and to periodically update this information, 2) to identify incident cases of cancers as well as deaths, and 3) to evaluate dietary and other risk factors for cancer incidence in a prospective study of men and women large enough to study important specific cancers within a reasonable follow-up period. We estimated that a cohort size of at least 80,000 men and 80,000 women would provide excellent statistical power (> 90%) within 5 years of follow-up to detect rate ratios as low as 1.3 for associations between incident breast carcinoma (1500 expected incidents) or prostate carcinoma (2500 expected incidents) and exposure prevalence of at least 10%.[53] Rate ratios of 1.5 could be detected with at least 90% power in gender specific analyses of colon carcinoma (600–900 expected incidents).[53]

In 1992 and 1993, a baseline questionnaire (described below) was mailed to 516,671 CPS-II men and women, ages 50–74 years, who resided in 21 states with population-based state cancer registries reported to ascertain at least 90% of incident cancer cases.[54] The states included California, Connecticut, Florida, Georgia, Illinois, Iowa, Louisiana, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Jersey, New Mexico, New York, North Carolina, Pennsylvania, Utah, Virginia, Washington, and Wisconsin. These states were chosen originally to enable follow-up of cancer incidence using computerized linkage with state cancer registries.

Development of the 1992–1993 questionnaire

The dietary assessment instrument used initially in the CPS-II Nutrition Cohort was a self-administered food frequency questionnaire (FFQ) that addressed usual eating habits over the past year. The questionnaire was a modification of the brief 60-item Health Habits and History Questionnaire (HHHQ) developed by Block and colleagues at the National Cancer Institute.[55] The brief HHHQ food list was developed based on the contributions of different foods to total population intake of energy and 17 other nutrients in the Second National Health and Nutrition Examination Survey (NHANES II) data and represents over 80% of dietary contribution to those intakes in the United States population.[55] Three portion sizes (small, medium, and large) are associated with each food item and were derived from gender and age specific portion size distributions in the NHANES II. In addition to portion size, the HHHQ assesses frequency of consumption for each item. Eight food items derived from the full version of the HHHQ[56] were added to the brief HHHQ for the Nutrition Cohort: other fruits, other vegetables, fats on vegetables, biscuits/muffins, yogurt, low-fat yogurt, tuna, and noodles/pasta. Additional questions derived from the full HHHQ or from the semiquantitative FFQ developed by Willett and colleagues[3,57] included changes in diet since 1982, degree to which red meat typically was cooked, and frequency of consumption of restaurant foods, low-fat foods, fats, and red meat. Questions pertaining to use of vitamin and mineral supplements also were included. In addition to the large section on dietary assessment, the questionnaire included briefer sections on demographics, medical history, weight history, smoking history, physical activity, medication use, occupational history, and, for women, breast and cervical screening, hormone replacement therapy, and family history of breast carcinoma.

Data collection and response

In the fall of 1992, the 10-page self-administered questionnaire was sent to 159,716 CPS-II men and their 140,780 CPS-II spouses ages 50–69 years and presumed to be living in the 21 states listed above. In response to this single mailing, we received completed, usable questionnaires from 55,105 CPS-II men and 50,937 CPS-II women who were still resident in the 21 states. To assemble both male and female cohorts of at least 80,000 individuals, we remailed the questionnaire in the fall of 1993. This mailing again was restricted to residents of the 21 selected states and was sent to all nonrespondents from the first mailing, to an additional 66,371 men ages 70–74 years and an additional 149,804 women ages 70–74 years or women ages 50–69 years who had not been included in 1992 because they were not living with a CPS-II male spouse. With this additional mailing, we received completed, usable questionnaires from a total of 86,406 men and 97,788 women ages 50–74 years (Table 1).

Validity and reproducibility of the FFQ

The FFQ used in the CPS-II Nutrition Cohort was validated in a stratified random sample of 244 men (mean age, 62 years) and 197 women (mean age, 60 years) who completed the survey prior to August 1993,

were cancer free, and did not report significant changes in diet or weight in the year prior to survey completion.[58] The validation study consisted of four 24-hour dietary recalls obtained by telephone for each person in the sample; telephone interviews were conducted from January 1994 to May 1995.

For total energy and most nutrients, intakes assessed by the Nutrition Cohort FFQ were lower than those from the 24-hour dietary recalls. Validity correlation coefficients for 10 food groups ranged from 0.1 for fats, oil, sweets, and snacks in women to 0.8 for alcoholic beverages in men, with a median correlation for all 10 food groups of 0.6 for both men and women. Validity coefficients for grams of energy-related macronutrients (protein, carbohydrate, ethanol, dietary fiber, and total and saturated fats) ranged from 0.3 (protein in men) to 0.8 (ethanol in men) and median values also were 0.6 for each gender group. Lower median validity correlations were observed for eight dietary vitamins and minerals: 0.4 in men and 0.3 in women. Median values for these micronutrients were influenced particularly by low correlation coefficients for zinc (< 0.2 in both men and women) and vitamin E (< 0.3 in both men and women). Higher correlations were seen for vitamin C (0.7 in men and women), calcium (0.6 in men and 0.7 in women) and folate (0.5 in men and 0.4 in women). Estimates for micronutrients from both the FFQ and the 24-hour recalls were from dietary sources only, exclusive of supplement use. Median values for reproducibility, based on a repeat administration of the FFQ, exceeded 0.7 for food groups, macronutrients and fats, and micronutrients among both men and women.

Nutrient data base

In preparation for nutritional analysis of the CPS-II Nutrition Cohort data, we estimated nutrient and food group intakes for each participant using the Dietary Analysis System (version 3.8a; Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, MD).[56] Modifications were made to the standard analysis configurations to account for differences between the FFQ used in the CPS-II Nutrition Cohort and the standard brief version of the HHHQ. We supplemented the nutrient data base with additional nutrient measures of interest, including glycemic load,[59] vitamin D,[60] and ethanol.[61]

Repeated exposure measurement

When the Nutrition Cohort originally was assembled in 1992 and 1993, we did not have funding to administer future questionnaires. Approval and funding for periodic resurveys were received in 1996, and we currently are recontacting the Nutrition Cohort members every 2 years using a mailed, self-administered questionnaire.

Accordingly, a new questionnaire was mailed to the 176,537 living members of the Nutrition Cohort beginning in September 1997. This survey included questions on personal and family history of cancer, screening history (for breast, colorectal, cervical, and prostate carcinoma), smoking status, physical activity, weight, waist size, hormone replacement therapy, vitamins, and medications. Close to 91% of living cohort members completed and returned this questionnaire ($n = 160{,}403$ cohort members). To achieve this level of response, we mailed the questionnaire six times at 3–4-month intervals. The first four mailings sent the complete questionnaire, and the next two mailings sent a much shorter version of the questionnaire that included questions on personal history of cancer and a few key exposure variables. We used certified mail for the final mailing. Finally, staff members at the ACS Call Center in Austin, Texas made telephone calls to remaining nonresponders and administered a short phone survey that was identical in content to the short mailed questionnaire. The entire process of mailings and phone calls was completed by August 1999. Ninety percent of the 160,403 respondents completed the long mailed questionnaire, 8% completed the short mailed questionnaire, and 2% completed the short phone questionnaire.

Another questionnaire was sent to the 166,773 living cohort members beginning in September 1999 using the mailing strategy outlined above. This questionnaire was devoted primarily to the administration of a second FFQ to obtain a more recent and complete assessment of diet. This FFQ is based on the 130-item FFQ developed by Willett and colleagues in 1986.[62,63] Just over 90% of living cohort members ($n = 150{,}362$ cohort members) completed any version of this questionnaire; 131,278 cohort members completed the long version, including the FFQ. The 2001 questionnaire repeats many of the questions asked in 1997 and includes new topics, such as herbal supplement use, sun exposure, and limitations in activities of daily living.

All questionnaires for the Nutrition Cohort have been designed for self-administration and for processing by optical scanning. Mailing, receiving, and scanning of the questionnaires have been conducted by outside contractors with the capacity for high-volume data processing. All questionnaires can be viewed on the ACS website at www.cancer.org.

Cancer incidence follow-up

We are capturing information about incident cancers by requesting this information in the periodic ques-

tionnaires to the study cohort and then validating reported cases of cancer through medical records acquisition. Cohort members provide written consent for medical records release. Pilot work linking the Nutrition Cohort members to four state cancer registries has indicated that the ability of our respondents to report a past diagnosis of cancer accurately is very high (sensitivity = 0.93 for the report of any cancer).[64] False positive reports were rare (specificity > 0.99). Additional work with 11 more registries yielded a range in sensitivity of 0.92–0.97 for a match on any cancer reported by the registry (data not published). Thus, we were confident that a system that validates self-reports of cancer in this cohort could acquire additional information on up to 92–97% of all cancers without incurring the large costs potentially associated with attempting to verify many false reports. This pilot work suggested that verifying participant reports of cancer would provide a timely and uniform method of ascertaining new (incident) cases. Verification includes checking for confirmation of any cancer diagnosis; it is not restricted to the specific disease site that was reported by the participant.

An incident cancer is defined as a cancer (other than nonmelanoma skin carcinoma) that reportedly occurred during the interval since the previous questionnaire or that was recorded on a death certificate during the interval since the previous questionnaire among cohort members who did not report the cancer at the beginning of the interval. Respondents to the 1997 questionnaire reported 9456 incident cancers in the interval between their date of enrollment in 1992 or 1993 and August 31, 1997 (Table 2). We have obtained medical records for 64% of these cases. In addition, 2485 deaths from cancer (as recorded on the death certificate) occurred in the same interval among cohort members who did not report the cancer at enrollment (Table 2). We are linking with state cancer registries as needed for follow-up of specific reported cases of cancer or cancer deaths when medical records are not available. Through this mechanism, we have obtained registry records for an additional 20% of reported cancer cases and for 74% of the interval cancer deaths. Verification of incident cancers currently is ongoing for cancers reported during the interval from 1997 to 1999.

TABLE 2
Incident Cancers in the Cancer Prevention Study II Nutrition Cohort, 1992/1993 to August 31, 1997

	Male		Female		
Site	Reported[a]	Deaths[b]	Reported[a]	Deaths[b]	Total
Prostate	3077	81	0	0	3158
Colon	524	120	378	71	1093
Lung	220	512	177	224	1133
Breast	16	0	1683	66	1765
Uterus	0	0	351	33	384
Ovary	0	0	140	72	212
Bladder	351	29	97	9	486
Kidney	99	45	34	19	197
Melanoma	735	26	435	12	1208
NHL	156	82	132	41	411
Leukemia	86	75	61	36	258
Other	376	597	328	335	1636
Total	5640	1567	3816	918	11,941

NHL: Non-Hodgkin's lymphoma.
[a] Reported on the 1997 questionnaire by cohort members who did not report the cancer in 1992/1993.
[b] Recorded on a death certificate through August 31, 1997 for cohort members who did not report the cancer in 1992/1993.

Collection of blood samples: The LifeLink Cohort

The CPS-II LifeLink Cohort was initiated in 1998 with the objective of collecting and storing blood samples from 40,000 members of the Nutrition Cohort. These donors were recruited from the approximately 142,000 surviving members of the cohort who lived in urban and suburban areas. Cohort members in more rural areas were not recruited for logistic reasons. Cohort members received an invitation by mail (nonrespondents were sent a second similar letter), and those who mailed back a form indicating their willingness to participate were telephoned to schedule a blood draw appointment at a participating hospital in their community. Blood sample collection was coordinated by ACS staff and volunteers and was performed by hospital staff at 312 community hospitals in 20 states (blood samples were not collected in Louisiana, because potential participants were dispersed geographically throughout the state). Hospitals were recruited from those approved by the American College of Surgeons (ACOS) Commission on Cancer with the assistance of the ACOS. Collection of blood samples was subject to Institutional Review Board approval in individual hospitals. Collection of blood samples for LifeLink was completed in June 2001. A total of 39,380 Nutrition Cohort members gave a single blood sample. In addition, we collected a second blood sample from 320 individuals to assess within-person variability in levels of blood analytes over time.

We collected a maximum of 43 mL of nonfasting whole blood from each participant drawn in two 15-mL tubes containing the anticoagulant ethylenediamine tetraacetic acid (EDTA) and a 13-mL serum separator tube (Fig. 1). The two 15-mL EDTA tubes yielded about 12 mL of plasma and 4 mL of red blood cells (additional red blood cell volume was discarded) as well as the buffy coat fraction, from which at least

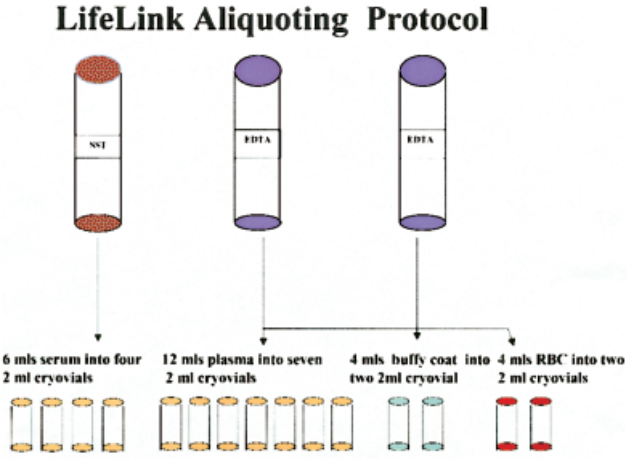

FIGURE 1. Diagram of the LifeLink aliquoting protocol.

340 μg of DNA can be extracted.[65] Samples were chilled on ice or stored in a refrigerator during blood collection. Hospital staff centrifuged the serum separator tubes to separate serum from cellular blood components prior to shipping, and all samples were then shipped in foam containers with coolant packs overnight by Express Mail to a central repository in Rockville, Maryland. Shipments of blood from the field were received, unpacked, aliquoted, and frozen in liquid nitrogen vapor phase at approximately −130 °C for long-term storage. All samples at the central repository are identified by bar code only.

We collected 450 mL of blood from each of 28 volunteer donors to create several different quality control (QC) *pools*. These will be used to evaluate the reliability of measurements made by collaborating laboratories, both before beginning an analysis and on an ongoing basis during all analyses. Specifically, one or more pair(s) of blinded identical samples from these QC pools will be interspersed with each batch of participant samples to be analyzed, and the concordance of results from these identical QC samples will be monitored. Aliquots from the QC pools are stored in the same liquid nitrogen freezers as the participant samples.

Collection of additional DNA from buccal cells.
In January 2001, we began an initiative to collect DNA from cohort members who were unable or unwilling to give a blood sample or who were not invited to give blood because they lived in a rural (i.e., geographically dispersed) area. Collection of buccal (cheek) cells is a quick, cost-effective, and noninvasive alternative means to obtain DNA. All living members of the Nutrition Cohort who did not give a blood sample (approximately 125,000 individuals) are being invited to provide a cheek cell sample by mail. Participants receive a sample collection kit that includes mouthwash, collection cups, instructions, and a postage-paid return envelope. They are instructed to swish the mouthwash in their mouths for 30 seconds, spit the sample into the cup, and return the sample in the envelope provided. This mouthwash method will yield approximately 30 μg of high-quality human DNA per participant.[66] Pilot work and early collection efforts have indicated that we will be able to collect a DNA sample from an additional 50,000 cohort members using this method. We anticipate that collection of buccal cell samples will be complete by mid-2002. Buccal cell DNA will be stored with the blood sample components at the central repository and will be identified by bar code only.

Protection of participants.
All aspects of the CPS-II protocol have been reviewed and approved by the Emory University School of Medicine Human Investigations Committee. Signed informed consent is received from participants to obtain medical records and from participants who donate blood or buccal cell samples. At the time of each mailed questionnaire, cohort members are informed that we use their identifying information to link with cancer registries and death indexes. In addition, in accordance with the provisions of section 301(d) of the Public Health Service Act (42 U.S. C. 241[d]), a Confidentiality Certificate has been issued for the CPS-II Nutrition Cohort by the National Cancer Institute. This certificate provides additional protection to study participants by helping ACS researchers avoid involuntary disclosures of participants' identifying information, which could expose them to adverse economic, psychological, and social consequences.

Characteristics of Nutrition Cohort Members

Because the majority of our prospective analyses will be conducted among individuals who were cancer free at baseline in 1992–1993, we present characteristics of cohort members after excluding 9358 men and 12,428 women who reported a history of cancer (other than nonmelanoma skin cancer). At baseline, virtually all members of this cohort were between ages 50 years and 79 years, with $>$ 50% between ages 60 years and 69 years (Table 3). The cohort is largely white (97%), married (89%), and has a relatively high level of education (38% are college graduates). Many cohort members are former smokers, especially among the men (60%), and relatively few cohort members were current smokers at baseline (7%). The prevalence of vitamin use (any type, at least once per week) was 38% in men and 52% in women. Over half of the cohort

TABLE 3
Baseline Characteristics of Persons with no Previous History of Cancer in the Cancer Prevention Study II Nutrition Cohort, 1992–1993[a]

Characteristic	Total %[b] (*n* = 162,408)	Men %[b] (*n* = 77,048)	Women %[b] (*n* = 85,360)
Age in 1992 (yrs)			
40–49	1.2	0.1	2.2
50–59	29.6	25.1	33.6
60–69	53.1	57.8	48.9
70–79	15.7	16.2	15.2
80+	0.4	0.8	0.1
Race			
White	97.3	97.3	97.2
Black	1.4	1.3	1.6
Other	1.3	1.4	1.2
Education			
Less than high school graduate	6.7	8.5	5.2
High school graduate	31.9	25.4	37.7
Some college	22.8	19.7	25.6
College graduate or more	37.9	45.7	30.9
Marital status			
Married	88.5	93.7	83.8
Widowed, separated, divorced	9.0	4.0	13.6
Never married	1.5	1.1	1.9
Smoking history			
Never smoker	44.3	32.6	54.7
Former smoker	47.9	59.9	37.1
Current smoker	6.6	6.6	6.6
Body mass index			
< 20	4.4	1.7	6.9
20 to < 25	39.6	33.8	44.9
25 to < 30	39.5	48.8	31.0
30 to < 35	11.4	11.8	11.0
≥ 35	3.5	2.4	4.5
Vitamin use			
No	45.0	52.2	38.6
Yes	45.2	38.2	51.5
Missing	9.8	9.6	10.0
Physical activity (in MET-hours per week)[c]			
None	10.7	12.3	9.3
> 0 to ≤ 7	31.7	29.4	33.7
> 7 to ≤ 17.5	29.8	28.1	31.4
> 17.5 to ≤ 31.5	19.9	21.8	18.1
> 31.5	6.3	6.8	5.8
Hormone replacement therapy			
Never use	—	—	49.8
Former user	—	—	13.1
Current user	—	—	32.8
Ever, unknown	—	—	1.4
High blood pressure	34.3	36.7	32.1
Blood pressure lowering drugs	26.6	28.7	24.6
Diabetes	7.0	8.6	5.6
High cholesterol	36.0	33.5	38.3
Cholesterol lowering drugs	9.4	10.2	8.6

MET: metabolic equivalent.

[a] Excludes persons with prevalent cancer, except nonmelanoma skin cancer (n = 9358 men; n = 12,428 women).

[b] Percentages that do not add up to the total reflect missing data.

[c] One hour of moderate paced walking = 3.5 MET-hours.

reported body weights that are classified as overweight (40% with a body mass index 25.0–29.9) or obese (15% with a body mass index ≥ 30.0) by World Health Organization criteria.[67]

Participants were asked to report the average time per week that they had spent doing seven leisure-time physical activities during the past year (walking, jogging/running, lap swimming, tennis or racquetball, bicycling/stationary bike, aerobics/calisthenics, and dancing). For the great majority of people, walking was the main type of physical activity. The median combined weekly energy expenditure from all of these activities was equivalent to between 2 and 3 hours of moderate walking for both men and women.[68]

Approximately 33% of women reported current use of hormone replacement therapy at baseline. Over 30% of both men and women reported physician-diagnosed high blood pressure, and over 75% of those reporting this condition reported taking blood pressure-lowering medication. Over 30% of both men and women also reported physician-diagnosed elevated cholesterol, but only one-third of these reported taking cholesterol-lowering drugs. Over 10% of men and 8% of women reported a history of diabetes mellitus (Table 3).

Mean values of total energy and selected nutrients from dietary sources in men and women are presented in Tables 4 and 5. For these estimates, we excluded men and women with > 15% of FFQ responses left blank and those with implausible reported energy intake (< 800 or > 4200 kcals per day for men and < 500 or > 3500 kcals per day for women). There appears to be a substantial range in intake of all nutrients in both men and women, as evidenced by the 25th and 75th percentile levels (Tables 4 and 5). For certain carotenoids, substantial differences in intake were observed by racial group (Table 6). Black men and women had considerably higher intakes of beta carotene (from sweet potatoes, collards and greens, and carrots), cryptoxanthin (from cornbread, corn, and peas), and lutein (from collards and greens and spinach) and considerably lower intake of lycopene (from tomatoes and strawberries) compared with white men and women.

TABLE 4
Mean and Range of Daily Intake of Selected Nutrients in Men, Cancer Prevention Study II Nutrition Cohort, 1992–1993

	Men (N = 69,216)[a]		
Nutrient	Mean	25th Percentile	75th percentile
Total calories (Kcal/day)	1805	1353	2164
Percent of calories from fat	35.4	29.7	41.4
Percent of calories from carbohydrates	44.7	39.0	50.0
Percent of calories from protein	16.1	14.0	17.9
Total fat (g)	72.6	47.0	91.7
Saturated fat (g)	23.9	14.7	30.5
Cholesterol (mg)	245	151	302
Carbohydrate (g)	198	149	238
Dietary fiber (g)	13.4	9.4	16.4
Protein (g)	71.4	53.0	85.8
Alpha carotene (mcg)	440	184	5560
Beta carotene (mcg)	2961	1510	3722
Calcium (mg)	846	534	1030
Cryptoxanthin (mcg)	69.7	24.2	93.5
Iron (mg)	12.0	8.87	14.5
Folate (mcg)	284	208	346
Linoleic acid (g)	13.9	8.6	17.9
Lutein (mcg)	1998	711	2503
Lycopene (mcg)	1325	616	1729
Magnesium (mg)	255	189	308
Niacin (mg)	17.8	13.3	21
Oleic acid (g)	27.0	17.2	34.2
Phosphorus (mg)	1245	901	1504
Potassium (mg)	2648	1996	3165
Retinol (mcg)	676	414	851
Riboflavin (B2) (mg)	1.87	1.32	2.27
Sodium (mg)	2860	2094	3458
Thiamin (B1) (mg)	1.35	1.02	1.62
Vitamin A (IU)	7863	4874	9685
Vitamin A (re)	1235	817	1523
Vitamin B6 (mg)	1.55	1.17	1.87
Vitamin C (mg)	123	76.2	156
Vitamin E (α-te)	9.27	6.60	11.4
Zinc (mg)	10.1	7.44	12.2
Vitamin D (IU)	190	107	235
Glycemic load	163	118	197
Ethanol (g)	11.2	0.00	13.6

[a] Exclusions include prevalent cancer, except nonmelanoma skin cancer; men with > 15% of the Food Frequency Questionnaire missing; and men with < 800 or > 4200 Kcals/day.

RESULTS AND DISCUSSION

The Nutrition Cohort was recruited from within the larger CPS-II Cohort. One great advantage of assembling the Nutrition Cohort in this way, rather than establishing a cohort of newly identified individuals, is the availability of historic measures of exposure collected in 1982. For this subcohort, the completion of the questionnaire in 1992–1993 established a new baseline for prospective analyses of cancer outcomes. Unlike previous ACS cohorts that were restricted to mortality follow-up, the Nutrition Cohort will include follow-up of incident cases of cancer, enhancing our ability to study predictors of both cancer occurrence and survival.

It is also important to note that the Nutrition Cohort is not intended to be a representative sample of the larger CPS-II Cohort. In recruiting the Nutrition Cohort, we did not attempt to maximize the participation rate but, rather, to enroll a selected cohort of a given size. Assembling a selected cohort rather than a

TABLE 5
Mean and Range of Daily Intake of Selected Nutrients in Women, Cancer Prevention Study II Nutrition Cohort, 1992–1993

	Women (N = 77,222)[a]		
Nutrient	Mean	25th Percentile	75th Percentile
Total calories (Kcal/day)	1360	1016	1629
Percent of calories from fat	33.8	27.1	40.5
Percent of calories from carbohydrates	47.5	41.5	53.4
Percent of calories from protein	16.7	14.5	18.8
Total fat (g)	52.2	32.8	66.1
Saturated fat (g)	16.5	9.9	20.9
Cholesterol (mg)	169	105	209
Carbohydrate (g)	160	118	194
Dietary fiber (g)	11.6	8.01	14.4
Protein (g)	56.1	41.3	67.8
Alpha carotene (mcg)	328	131	421
Beta carotene (mcg)	2309	1185	2901
Calcium (mg)	752	456	931
Cryptoxanthin (mcg)	63.1	21.6	87.3
Iron (mg)	9.81	6.94	12.1
Folate (mcg)	253	169	312
Linoleic acid (g)	10.3	5.76	13.4
Lutein (mcg)	1710	643	2073
Lycopene (mcg)	1123	466	1458
Magnesium (mg)	210	152	257
Niacin (mg)	14.3	10.2	17.3
Oleic acid (g)	19.3	11.9	25.7
Phosphorus (mg)	938	707	1237
Potassium (mg)	2187	1598	2652
Retinol (mcg)	551	332	707
Riboflavin (B2) (mg)	1.64	1.07	2.05
Sodium (mg)	2205	1590	2682
Thiamin (B1) (mg)	1.16	0.82	1.41
Vitamin A (IU)	6276	3899	7775
Vitamin A (re)	993	653	1236
Vitamin B6 (mg)	1.31	0.91	1.59
Vitamin C (mg)	119	72.5	153
Vitamin E (α-te)	8.69	5.29	10.5
Zinc (mg)	8.51	5.73	10.3
Vitamin D (IU)	167	87.5	213
Glycemic load	126	90.8	154
Ethanol (g)	4.70	0.00	4.71

[a] Exclusions include prevalent cancer, except nonmelanoma skin cancer; women with > 15% of the Food Frequency Questionnaire missing; and women with < 500 or > 3500 Kcals/day.

TABLE 6
Mean and Range of Daily Intake of Selected Nutrients by Gender and Race, Cancer Prevention Study II Nutrition Cohort, 1992–1993

	Men[a]		Women[a]	
Nutrient	White (N = 67,541)	Black (N = 750)	White (N = 75,270)	Black (N = 1043)
Total calories (Kcal/day)	1807	1731	1361	1335
Percent of calories from fat	35.4	36.1	33.8	34.8
Percent of calories from carbohydrates	44.6	45.1	47.5	47.8
Percent of calories from protein	16.1	15.6	16.7	16.3
Total fat (g)	72.8	70.3	52.3	53.0
Saturated fat (g)	23.9	23.1	16.5	16.7
Cholesterol (mg)	245	266	168	183
Carbohydrate (g)	199	193	160	157
Dietary fiber (g)	13.4	12.7	11.6	11.2
Protein (g)	71.5	67.0	56.2	53.7
Alpha carotene (mcg)	441	341	329	271
Beta carotene (mcg)	2945	4468	2291	3680
Calcium (mg)	849	724	755	656
Cryptoxanthin (mcg)	69.3	95.5	62.7	83.3
Iron (mg)	12.0	11.3	9.82	9.51
Folate (mcg)	285	269	253	252
Linoleic acid (g)	13.9	13.2	10.3	10.4
Lutein (mcg)	1961	4950	1669	4323
Lycopene (mcg)	1332	963	1129	827
Magnesium (mg)	256	239	210	202
Niacin (mg)	17.8	16.8	14.4	14.2
Oleic acid (g)	27.1	26.6	19.3	19.9
Phosphorus (mg)	1248	1128	1014	931
Potassium (mg)	2655	2385	2193	2045
Retinol (mcg)	677	689	552	582
Riboflavin (B2) (mg)	1.88	1.71	1.64	1.54
Sodium (mg)	2863	2754	2207	2122
Thiamin (B1) (mg)	1.35	1.31	1.16	1.15
Vitamin A (IU)	7865	8543	6273	7130
Vitamin A (re)	1236	1313	994	1100
Vitamin B6 (mg)	1.59	1.43	1.32	1.27
Vitamin C (mg)	123	127	119	130
Vitamin E (α-te)	9.29	8.83	870	8.81
Zinc (mg)	10.2	9.19	8.53	8.05
Vitamin D (IU)	191	164	167	148
Glycemic load	163	160	126	124
Ethanol (g)	11.3	8.78	4.75	2.67

[a] Exclusions include prevalent cancer, except nonmelanoma skin cancer; persons with > 15% of the Food Frequency Questionnaire missing; men with < 800 or > 4200 Kcal/day; and women with < 500 or > 3500 Kcal/day.

population-based cohort has several advantages. Most noteworthy, cohort stability and ongoing, long-term participation that are often seen in selected cohorts[69–72] increase the likelihood of achieving high follow-up rates essential to the internal validity of exposure-disease results from a cohort study.[2,73] In addition, a selected cohort usually offers reduced potential for confounding and may yield higher data quality, depending on the nature of the cohort.[2] Finally, these factors all contribute to greatly increased cost effectiveness because of the relative ease of tracing and following cohort members.[73]

An often stated concern about selected cohorts is the potential loss of generalizability (external validity). To address this concern, it is critical to differentiate between the generalizability of an absolute measure of disease (or risk factor) frequency and the generaliz-

ability of measures of association. Absolute rates of disease or risk factor frequency that are calculated in a selected cohort probably will not be generalizable to a different population. However, the purpose of most large cohort studies is to study relations (associations) between exposure and disease. If the internal validity of the study is maintained, then measures of association typically are unbiased and are generalizable even if the cohort is not a random sample of a defined underlying population.[2,73–76] For example, although the British Doctors Study[77,78] was not a random sample of British physicians, let alone the general population, the associations between smoking and disease observed in that study have proved to be generalizable to many other populations. Moreover, generalizability of associations across different subgroups within a cohort can be examined carefully and quantified.

The brief FFQ that was used to capture dietary intake in 1992–1993 underestimated intake of total energy and most individual nutrients. However, if this underestimation is fairly uniform across cohort members, as we expect it is, then it should not invalidate internal comparisons between high-consuming and low-consuming groups. Also, the range of dietary intake appears sufficient to allow detection of true differences. Although median validity correlation coefficients for food groups, macronutrients, and many micronutrients were comparable to those observed in other studies,[57,62,79] validity coefficients were low for some micronutrients (e.g., zinc and vitamin E) from dietary sources. Many of these coefficients are likely to improve when vitamin supplement use is considered. In addition, the 1999 questionnaire recaptured dietary intake in this cohort using a more detailed FFQ[62,63] than was used in 1992. The 1992 FFQ will be used to define dietary exposures for those nutrients measured with reasonable validity and for cases of cancer occurring from 1992 to 1999. The 1999 FFQ will provide the primary dietary exposure data for all cases of cancer occurring after 1999. With rare exception, any cumulative measures of exposure estimated using both FFQs will be based on quantile ordering of individuals and not on absolute measures of consumption, given the inherent differences in the two instruments.

Blood and buccal cell samples will be used by ACS researchers and collaborating investigators for future nested case–control studies on the correlations between hormonal, dietary, genetic, and other factors and important health outcomes, particularly breast, prostate, and colorectal carcinomas. To oversee this valuable resource, ACS will form a standing review panel comprised of intramural and extramural scientists, including epidemiologists familiar with large cohort studies and laboratory and other researchers able to integrate cancer epidemiology, biology, genetics, and laboratory research. We also are participating in a National Cancer Institute initiative to form a Consortium of Cohort Studies sufficiently large to address interactions between genetic factors and environmental exposures. No single study is currently large enough for this purpose, because, typically, several thousand cases are needed to provide statistically stable estimates of gene-environment or gene-gene interactions.

The ACS Nutrition Cohort establishes a new, large, prospective cohort of both men and women, with detailed information on their dietary and other exposure characteristics, who will be followed actively for all new cases of cancer and for all causes of death. This cohort promises to be particularly valuable for the study of cancer occurrence, mortality, and survival as they relate to obesity and weight change, physical activity at various points in life, vitamin supplement use, exogenous hormone use, other medications (such as aspirin and nonsteroidal anti-inflammatory drugs), and cancer screening modalities. In addition to the ongoing collection of exposure and disease information in the future, this cohort benefits from the availability of exposure data collected 10 years before baseline, in 1982, providing a unique opportunity to study temporal changes in exposure and the subsequent impact on the occurrence of disease.

REFERENCES

1. Samet J, Munoz A. Evolution of the cohort study. *Epidemiol Rev* 1998;20:1–14.
2. Willett W, Colditz G. Approaches for conducting large cohort studies. *Epidemiol Rev* 1998;20:91–9.
3. Willett W. Nutritional epidemiology, 2nd ed. New York: Oxford University Press, 1998;484–96.
4. Riboli E, Kaaks R. The EPIC project: rationale and study design. *Int J Epidemiol* 1997;26(Suppl 1):S6–14.
5. Kolonel L, Henderson B, Hankin J, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;151:346–57.
6. Thun M, Calle E, Rodriguez C, Wingo P. Epidemiologic research at the American Cancer Society. *Cancer Epidemiol Biomarkers Prev* 2000;9:861–8.
7. Hammond E, Horn D. Smoking and death rates—report on 44 months of follow-up on 187,783 men. Part I. Total mortality. *JAMA* 1958;166:1159–72.
8. Hammond E, Horn D. Smoking and death rates—report on 44 months of follow-up on 187,783 men. Part II. Death rates by cause. *JAMA* 1958;166:1294–308.
9. Hammond C. Smoking in relation to the death rates of one million men and women. In: Haenszel W, editor. Epidemiologic approaches to the study of cancer and other chronic diseases. National Cancer Institute monograph 19. Bethesda: US Department of Health, Education, and Welfare, National Cancer Institute, 1966:124–204.
10. Thun M, Lally C, Calle E, Flanders W, Heath C. Excess mortality among cigarette smokers: changes in a 20-year interval. *Am J Public Health* 1995;85:1223–30.

11. Thun M, Heath C. Changes in mortality from smoking in two American Cancer Society prospective studies since 1959. *Prev Med* 1997;26:422–6.
12. Thun M, Lally C, Flannery J, Calle E, Flanders W, Heath C. Cigarette smoking and changes in the histopathology of lung cancer. *J Natl Cancer Inst* 1997;89:1580–6.
13. Lew E, Garfinkle L. Variations in mortality by weight among 750,000 men and women. J *Chron Dis* 1979;32:563–76.
14. Garfinkel L. Overweight and cancer. *Ann Intern Med* 1985; 103(6):1034–6.
15. Garfinkel L. Overweight and mortality. *Cancer* 1986;58(8): 1826–9.
16. Williamson D, Pamuk E, Thun M, Flander W, Byers T, Heath C. A prospective study of intentional weight loss and mortality in never smoking overweight US white women aged 40–64 years. *Am J Epidemiol* 1995;141:1128–41.
17. Stevens J, Plankey M, Williamson D, Thun MJ, Rust PF, Palesch Y, et al. The body mass index—mortality relationship in white and African American women. *Obesity Res* 1998;6:268–77.
18. Stevens J, Cai J, Pamuk E, Williamson D, Thun M, Wood J. The effect of age on the association between body-mass index and mortality. *N Engl J Med* 1998;338:1–7.
19. Hammond E. Smoking habits and air pollution in relation to lung cancer. In: Lee D, editor. Environmental factors in respiratory disease. New York: Academic Press, 1972.
20. Holmberg L, Ekbom A, Calle E, Mokdad A, Byers T. Parental age and breast cancer mortality: a cohort study of 384,000 women. *Epidemiology* 1995;6:425–7.
21. Michels-Blanck H, Byers T, Mokdad A, Will J, Calle E. Menstrual patterns and breast cancer mortality. *Epidemiology* 1996;7:543–6.
22. Holmberg L, Ekbom A, Calle E, Mokdad A, Byers T. Breast cancer mortality in relation to self-reported use of breast self-examination: a cohort study of 450,000 women. *Breast Cancer Res Treat* 1997;43:137–40.
23. Garfinkel L. Selection, follow-up, and analysis in the American Cancer Society prospective studies. *Natl Cancer Inst Mongr* 1985;67:49–52.
24. Calle EE, Terrell DD. Utility of the National Death Index for ascertainment of mortality among Cancer Prevention Study II participants. *Am J Epidemiol* 1993;137(2):235–41.
25. Thun M, Peto R, Lopez A, Monaco JH, Henley SJ, Heath CW Jr., et al. Alcohol consumption and mortality in middle-aged and elderly US adults. *N Engl J Med* 1997;337:1705–14.
26. Rodriguez C, Calle E, Coates R, Miracle-McMahill H, Thun M, Heath C. Estrogen replacement therapy and fatal ovarian cancer. *Am J Epidemiol* 1995;141:828–35.
27. Calle E, Miracle-McMahill H, Thun M, Heath C. Estrogen replacement therapy and risk of fatal colon cancer in a prospective cohort of postmenopausal women. *J Natl Cancer Inst* 1995;87:517–23.
28. Willis D, Calle E, Miracle-McMahill H, Heath C. Estrogen replacement therapy and risk of fatal breast cancer in a prospective cohort of postmenopausal women in the United States. *Cancer Causes Control* 1996;7:449–57.
29. Calle E, Mervis C, Thun M, Rodriguez C, Wingo P, Heath C. Diethylstilbestrol and risk of fatal breast cancer. *Am J Epidemiol* 1996;144:645–52.
30. Rodriguez C, Patel A, Calle E, Thun M. Estrogen replacement therapy and ovarian cancer mortality in a large prospective study of U.S. women. *JAMA* 2001;285:1460–5.
31. Thun M, Namboodiri M, Heath C. Aspirin use and reduced risk of fatal colon cancer. *N Engl J Med* 1991;325:1593–6.
32. Thun M, Namboodiri M, Calle E, Flanders D, Heath C. Aspirin use and risk of fatal cancer. *Cancer Res* 1993;53: 1322–7.
33. Watkins M, Erickson J, Thun M, Mulinare J, Heath C. Multivitamin use and mortality in a large prospective study. *Am J Epidemiol* 2000;152:149–62.
34. Jacobs E, Patel A, Chao A, Rodriguez C, Calle E, Thun M. Vitamin C and vitamin E supplement use and colorectal cancer mortality in a large American Cancer Society cohort. *Cancer Epidemiol Biomarkers Prev* 2001;10:17–23.
35. Pope C III, Thun M, Namboodiri M, Dockery DW, Evans JS, Speizer FE, et al. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. *Am J Respir Crit Care Med* 1995;151:669–74.
36. Steenland K, Thun M, Lally C, Heath C. Environmental tobacco smoking and ischemic heart disease in the American Cancer Society CPS-II cohort. *Circulation* 1996;94: 622–8.
37. Cardenas V, Thun M, Austin H, Lally CA, Clark WS, Greenberg RS, et al. Environmental tobacco smoke and lung cancer mortality in the American Cancer Society's Cancer Prevention II Study. *Cancer Causes Control* 1997;8:57–64.
38. Thun MJ, Myers DG, Day-Lally C, Namboodiri MM, Calle E, Flanders WD et al. Age and the exposure-response relationships between cigarette smoking and premature death in Cancer Prevention Study II National Cancer Institute, Smoking and Tobacco Control, monograph 8: changes in cigarette-related disease risks and their implication for prevention and control. NIH publication no. 97-4213. Washington, DC: National Institutes of Health, 1997:383–475.
39. Jacobs E, Thun M, Apicella L. Cigar smoking and death from coronary heart disease in a prospective study of United States men. *Arch Intern Med* 1999;159:2413–8.
40. Thun M, Apicella L, Henley S. Estimating the number of smoking-related deaths. *JAMA* 2000;284:2319–20.
41. Wartenberg D, Calle E, Thun M, Heath C, Lally C, Woodruff T. Passive smoking exposure and female breast cancer mortality. *J Natl Cancer Inst* 2000;92:1666–73.
42. Shapiro J, Jacobs E, Thun M. Cigar smoking in men and risk of death from tobacco-related cancers. *J Natl Cancer Inst* 2000;92:333–7.
43. Chao A, Thun M, Jacobs E, Henley S, Rodriguez C, Calle E. Cigarette smoking and colorectal cancer mortality in Cancer Prevention Study II. *J Natl Cancer Inst* 2000;92:1888–96.
44. Calle E, Martin L, Thun M, Miracle H, Heath C. Family history, age, and risk of fatal breast cancer. *Am J Epidemiol* 1993;138:675–81.
45. Rodriguez C, Calle EE, Miracle-McMahill HL, Tatham LM, Wingo PA, Thun MJ, et al. Family history and risk of fatal prostate cancer. *Epidemiology* 1997;8:653–7.
46. Rodriguez C, Calle EE, Tatham LM, Wingo PA, Miracle-McMahill HL, Thun MJ, et al. Family history of breast cancer as a predictor for fatal prostate cancer. *Epidemiology* 1998; 9:525–9.
47. Boffetta P, Stellman S, Garfinkel L. Diesel exhaust exposure and mortality among males in the American Cancer Society prospective study. *Am J Ind Med* 1988;14:403–15.
48. Calle E, Murphy T, Rodriguez C, Thun M, Heath C. Occupation and breast cancer mortality in a prospective cohort of US women. *Am J Epidemiol* 1998;148:191–7.
49. Kahn H, Tatham L, Rodriguez C, Calle E, Thun M, Heath C. Stable behaviors associated with adults' 10-year change in body mass index and likelihood of gain at the waist. *Am J Public Health* 1997;87:747–54.

50. Calle E, Thun M, Petrelli J, Rodriguez C, Heath C. Body mass index and mortality in a prospective cohort of U.S. adults. *N Engl J Med* 1999;341:1097–105.
51. Murphy T, Calle E, Rodriguez C, Kahn H, Thun M. Body mass index and colon cancer mortality in a large prospective study. *Am J Epidemiol* 2000;152:847–54.
52. Rodriguez C, Patel A, Calle E, Jacobs E, Chao A, Thun M. BMI, height and prostate cancer mortality in two large cohorts of U.S. adult men. *Cancer Epidemiol Biomarkers Prev* 2001;10:345–53.
53. Schlesslman J. Case-control studies: design, conduct, analysis. New York: Oxford University Press, 1982.
54. Howe H. Cancer incidence in North America, 1988–1991. Sacramento: North American Associations of Central Cancer Registries, 1995.
55. Block G, Hartman A, Naughton D. A reduced dietary questionnaire: development and validation. *Epidemiology* 1990; 1:58–64.
56. Block G, Coyl L, Smucker R, Harlan L. Health Habits and History Questionnaire: diet history and other risk factors [personal computer system documentation]. Bethesda, MD: Division of Cancer Prevention and Control, National Cancer Institute, National Institutes of Health, 1989.
57. Willett W, Sampson L, Stampfer M, Rosner B, Bain C, Witschi J, et al. Reproducibility and validity of a semi-quantitative food frequency questionnaire. *Am J Epidemiol* 1985;122: 51–65.
58. Flagg E, Coates R, Calle E, Potischman N, Thun M. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey Cohort food frequency questionnaire. *Epidemiology* 2000;11:462–8.
59. Foster-Powell K, Brand-Miller J. International tables of glycemic index. *Am J Clin Nutr* 1995;62:S871–93.
60. U.S. Department of Agriculture. Provisional table on vitamin D content of foods. Beltsville, MD: USDA Nutrient Data Laboratory, 1991.
61. U.S. Department of Agriculture, Agricultural Research Service, 1999. USDA nutrient database for standard reference, Release 13. Nutrient data laboratory homepage http://www.nal.usda.gov/fnic/foodcomp.
62. Rimm E, Giovannucci E, Stampfer M, Colditz G, Litin L, Willett W. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. *Am J Epidemiol* 1992;135:1114–26.
63. Willett W. Reproducibility and validity of food-frequency questionnaires Nutritional Epidemiology. New York: Oxford University Press, 1998.
64. Bergmann M, Calle E, Mervis C, Miracle-McMahill H, Thun M, Heath C. Validity of self-reported cancers in a prospective cohort study in comparison to data from state cancer registries. *Am J Epidemiol* 1998;147:556–62.
65. Hankinson S, London S, Chute C, Barbieri RL, Jones L, Kaplan LA, et al. Effect of transport conditions on the stability of biochemical markers in blood. *Clin Chem* 1989;35: 2313–6.
66. Feigelson H, Rodriguez C, Robertson A, Jacobs E, Calle EE, Reid Y, et al. Determinants of DNA yield and quality from buccal cell samples collected with mouthwash. *Cancer Epidemiol Biomarkers Prev* 2001;10:1005–8.
67. World Health Organization. WHO Expert Committee on Physical Status. Physical status: the use and interpretation of anthropometry. Geneva: World Health Organization, 1995.
68. Ainsworth BE, Haskell WL, Leon AS, Jacobs DR, Montoye HJ, Sallis JF, Paffenbarger RS. Compendium of physical activities: classification of energy costs of human physical activities. *Med Sci Sports Exercise* 1993;25:71–80.
69. Doll R, Hill A. Lung cancer and other causes of death in relation to smoking: a second report on the mortality of British doctors. *Br Med J* 1956;2:1071–81.
70. Colditz G. The Nurses' Health Study: a cohort of US women followed since 1976. *J Am Med Womens Assoc* 1995;50:40–4.
71. Rimm E, Stampfer M, Ascherio A, Giovannucci E, Colditz G, Willett W. Dietary antioxidant intake and risk of coronary heart disease among men. *N Engl J Med* 1993;328:1450–6.
72. Fraser G, Beeson W, Phillips R. Diet and lung cancer risk in California Seventh Day Adventists. *Am J Epidemiol* 1991;133: 683–93.
73. Szklo M. Population-based cohort studies. *Epidemiol Rev* 1998;20:81–90.
74. Dawber T. The Framingham study: the epidemiology of atherosclerotic disease. Cambridge: Harvard University Press, 1980.
75. Keys A. Seven countries: a multivariate analysis of death and coronary heart disease. Cambridge: Harvard University Press, 1980.
76. Rothman KJ, Greenland S. Modern epidemiology. Philadelphia: Lippincott-Raven Publishers, 1998:133–4.
77. Doll R, Hill A. The mortality of doctors in relation to their smoking habits: a preliminary report. *Br Med J* 1954;1: 1451–5.
78. Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994;309:901–11.
79. Munger R, Folsom A, Kushi L, Kaye S, Sellers T. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. *Am J Epidemiol* 1992;136:192–200.