The objective of the proposed data analysis project is to assess whether high exposure to vitamin D from sunlight exposure, diet, and dietary supplements is associated with reduced breast cancer risk. We will perform a retrospective cohort analysis based on data provided by the cohort of women aged 25 to 74 years who participated in the first National Health and Nutrition Examination Survey (NHANES I) from 1971 to 1975 and who were followed-up in the NHANES I Epidemiologic Follow-Up Studies conducted in 1982-84, 1986, and 1987. The average follow-up period was 15 years, during which 190 newly diagnosed breast cancers occurred. For each vitamin D-related exposure measure, we will estimate the incidence of breast cancer among exposed women and unexposed women, and we will estimate the relative risk associated with these exposure variables using the Cox proportional hazards model and Poisson regression, adjusting for potentially confounding variables. This data analysis project is a cost-effective way to explore a new hypothesis addressing potentially modifiable risk factors for breast cancer.
The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to stay within the lines to meet optical scanning requirements.

Block 1. **Agency Use Only (Leave blank).**

Block 2. **Report Date.** Full publication date including day, month, and year, if available (e.g. 1 Jan 88). Must cite at least the year.

Block 3. **Type of Report and Dates Covered.** State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. **Title and Subtitle.** A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. **Funding Numbers.** To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

- C - Contract
- G - Grant
- PE - Program
- PR - Project
- TA - Task
- WU - Work Unit
- Accession No.

Block 6. **Author(s).** Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. **Performing Organization Name(s) and Address(es).** Self-explanatory.

Block 8. **Performing Organization Report Number.** Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. **Sponsoring/Monitoring Agency Name(s) and Address(es).** Self-explanatory.

Block 10. **Sponsoring/Monitoring Agency Report Number. (If known)**

Block 11. **Supplementary Notes.** Enter information not included elsewhere such as: Prepared in cooperation with...; Trans. of...; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. **Distribution/Availability Statement.** Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

- **DOD** - See DoDD 5230.24, “Distribution Statements on Technical Documents.”
- **DOE** - See authorities.
- **NTIS** - Leave blank.

Block 12b. **Distribution Code.**

- **DOD** - Leave blank.
- **DOE** - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.
- **NASA** - Leave blank.
- **NTIS** - Leave blank.

Block 13. **Abstract.** Include a brief (Maximum 200 words) factual summary of the most significant information contained in the report.

Block 14. **Subject Terms.** Keywords or phrases identifying major subjects in the report.

Block 15. **Number of Pages.** Enter the total number of pages.

Block 16. **Price Code.** Enter appropriate price code (NTIS only).

Blocks 17.-18. **Security Classifications.** Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. **Limitation of Abstract.** This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.
Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

Where copyrighted material is quoted, permission has been obtained to use such material.

Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

[Signature] 1-23-96

PI - Signature Date
# VITAMIN D AND BREAST CANCER

## ANNUAL REPORT

### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION</td>
<td>2</td>
</tr>
<tr>
<td>1.1. Nature of the problem</td>
<td>2</td>
</tr>
<tr>
<td>1.1. Background of previous work</td>
<td>2</td>
</tr>
<tr>
<td>1.2. Purpose of present work</td>
<td>3</td>
</tr>
<tr>
<td>1.3. Methods of approach</td>
<td>3</td>
</tr>
<tr>
<td>2. BODY</td>
<td>3</td>
</tr>
<tr>
<td>2.1. Methods used</td>
<td>3</td>
</tr>
<tr>
<td>2.2. Results</td>
<td>3</td>
</tr>
<tr>
<td>2.2.1. Extraction of exposure and confounder variables</td>
<td>4</td>
</tr>
<tr>
<td>2.2.2. Adding of vitamin D to NHANES I nutrient database</td>
<td>4</td>
</tr>
<tr>
<td>2.2.3. Identification of analytic cohort</td>
<td>6</td>
</tr>
<tr>
<td>2.2.4. Identification of breast cancer cases</td>
<td>6</td>
</tr>
<tr>
<td>2.2.5. Estimation of person-years of follow-up</td>
<td>7</td>
</tr>
<tr>
<td>2.3. Discussion</td>
<td>7</td>
</tr>
<tr>
<td>3. LITERATURE CITED</td>
<td>8</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

1.1. Nature of the problem

Breast cancer is the leading incident cancer in the United States, affecting one in nine women over their lifetimes, and accounting for 32% of all newly diagnosed cancers in women. Yet the etiology of breast cancer is not well understood. As recently summarized [1], the most consistently reported risk factors for breast cancer include menstrual and reproductive characteristics, such as early menarche, late age at first full-term pregnancy, low parity, and late age at menopause. Other established risk factors include high education, postmenopausal obesity, a family history of breast cancer, a personal history of benign breast disease, and ionizing radiation to the chest. These risk factors, however, account for less than half of the incidence of breast cancer [2, 3]. In addition, few of the established risk factors are potentially modifiable through behavioral or environmental changes. Epidemiologic research into new risk factors for breast cancer is clearly needed in order to prevent this important cause of morbidity and mortality. The ongoing study addresses the role of vitamin D, a newly hypothesized risk factor which is potentially modifiable.

1.2. Background of previous work

Breast cancer mortality rates for both black and white women are higher in the Northeast than in the South of the US [4]. Although the geographic variation has somewhat diminished over time, as more areas in the South have experienced rising mortality rates than in the North [5], state-level mortality rates in 1985 to 1989 were still about 50% higher in the Northeast than in the South [4].

Until recently, the north-south gradient of breast cancer mortality rates within the United States remained unexplained. A north-south gradient is not evident for most other cancers [6]. Therefore, the observed geographic variation is unlikely to be due solely to regional differences in death certification. An analysis of county-level breast cancer mortality rates found only weak correlations with income, level of urbanization, and birth rates among young women [7]. An ecologic correlation study reported an inverse association between breast cancer mortality rates and solar radiation, the major source of vitamin D [8]. Accordingly, Garland et al. hypothesized that vitamin D, which is synthesized by the skin following sunlight exposure and absorbed from the diet, may reduce breast cancer risk [8]. A new correlation study published in late 1995 reported that most of the differences in mortality rates between the North East and the South were explained by regional differences in reproductive risk factors [9]. The authors, however, concluded that regional differences in exposure to environmental factors such as vitamin D, sunlight exposures, pesticides etc., may account for the remaining geographic differences in mortality rates.

Although the vitamin D hypothesis was posed in 1990 [8], no published epidemiologic analytic study to date has directly tested the hypothesis that high serum levels of 1,25-dihydroxyvitamin D may protect against the development of breast cancer. The strongest evidence supporting the plausibility of the vitamin D hypothesis stems from experimental studies. Over the past 10 to 15 years experimental evidence has accumulated on the anti-cancer effects of vitamin D. Both in vitro and in vivo studies have
demonstrated that 1,25-dihydroxyvitamin D (1,25(OH)2D), the biologically active metabolite of vitamin D, inhibits the proliferation and promotes the differentiation of many types of normal and malignant cells, including breast cancer cells [10-12]. The action of 1,25(OH)2D is mediated through specific intracellular receptors that have been identified in many cell types [13, 14], including breast cancer cells [15]. A number of vitamin D analogs have recently been developed that also inhibit cell proliferation in vitro and in vivo, but with a fraction of the calcemic activity of 1,25(OH)2D (31, 32). Vitamin D analogues therefore may have potential future use in chemoprevention [16].

1.2. Purpose of present work

The purpose of the on-going study is to assess whether exposure to high levels of vitamin D is associated with reduced breast cancer risk. Associations with various measures of sunlight exposure, as well as vitamin D intake from diet and dietary supplements will be investigated. If high exposure to vitamin D indeed reduces breast cancer risk, the proposed study would make an important contribution towards the identification of potentially modifiable risk factors.

1.3. Methods of approach

The on-going study is testing the vitamin D hypothesis analyzing existing data from a national health survey. The investigator will perform a retrospective cohort analysis based on data provided by the cohort of women aged 25 to 74 years who participated in the first National Health and Nutrition Examination Survey (NHANES I) from 1971 to 1975 and who were followed-up in the NHANES I Epidemiologic Follow-up Studies (NHEFS) conducted in 1982-84, 1986, and 1987. The baseline and first follow-up interview collected information on several variables that relate to vitamin D, including sunlight exposure and intake of vitamin D from food and supplements. For each vitamin D-related exposure measure, the incidence of breast cancer among exposed women and unexposed women will be estimated. The relative risk associated with these exposure variables will be estimated using the Cox proportional hazards model and Poisson regression, adjusting for potentially confounding variables.

2. BODY

2.1. Methods used

In preparation for the conduct of the planned retrospective cohort analysis, a large database has been created for the cohort of women who were first interviewed and examined in 1971-75, and traced and re-interviewed in 1982-84, 1986, and 1987. Relevant data on exposure, confounder, and outcome variables were extracted from 14 NHANES data tapes and merged into a single database. The statistical analysis system (SAS) was used to conduct this data management task.

2.2. Results

The goals for the first year of the on-going study were five-fold: (1) to build a SAS database for the cohort of women included in the NHANES I Epidemiologic Follow-up
Study by extracting the relevant exposure and confounder variables from the various NHANES data tapes; (2) to add vitamin D to the NHANES I nutrient database; (3) to establish the analytic cohort of women who were traced and interviewed in the follow-up surveys conducted in 1982-84, 1986, and 1987; (4) to identify all women who were diagnosed with breast cancer and/or died from breast cancer during the follow-up period; and (5) to estimate the person-years of follow-up for each individual in the analytic cohort. Results achieved during the first year as they pertain to each of these goals are described below.

2.2.1. Extraction of exposure and confounder variables

The extraction of exposure and confounder variables from 6 data tapes and merging into a single SAS database has been completed. The current database includes (1) exposure variables which are direct or indirect measures of sunlight exposure (e.g., degree of actinic skin damage, frequency of usual job and leisure time-related outdoor activities, state of birth, residence of longest duration, region of residence at baseline, longest-held occupation, job held at baseline interview, sun exposure on job, sun exposure in leisure time, skin reaction to sun, natural hair color, eye color); (3) exposure variables which are direct or indirect measures of dietary vitamin D intake (e.g., frequency of consumption of vitamin D-rich foods such as dairy products, fish, eggs, avoidance of milk, avoidance of seafood); (3) exposure variables of vitamin D intake from supplements (e.g., single vitamin D supplements, multivitamins, cod liver oil); and (4) several variables on other risk factors which will be considered as potential confounders in the analysis (e.g., age, race/ethnicity, education, marital status, income, weight, height, total calorie intake, total fat intake, physical activity, alcohol intake, age at menarche, age at first birth, parity, family history of breast cancer, ). Additional data on the use of single vitamin D supplements which is not included in the public use data tapes was obtained from Dr. Lee Young at the NCI Department of Cancer Prevention.

2.2.2. Adding of vitamin D to NHANES I nutrient database

The NHANES I nutrient database does not include vitamin D. While trying to identify the best strategy to assign vitamin D nutrient values to the 3,527 food items reported in the 24-hour dietary recall by NHANES I participants, we learned that Dr. Suzanne Murphy at the University of California at Berkeley had added vitamin D nutrient values to the NHANES I nutrient database. For an analysis of dietary NHANES I data and cardiovascular disease, Dr. Murphy used a more current and complete nutrient database than the NHANES I nutrient database which was created in the 1970s and includes 18 nutrients only. The UC Berkeley Minilist which was developed in the 1970’s by Dr. Jean Pennington [21] and has been updated since then, includes data on 35 nutrients, including vitamin D, for 195 basic food ingredients. Dr. Murphy then cross-referenced the 3,527 food codes from NHANES I with the 195 food codes from the Minilist, using direct substitutions for simple foods, or combinations of Minilist food codes for mixtures. This methodology is described in an unpublished report [19] and a published abstract [20].

After reviewing Dr. Murphy’s methodology and meeting with her, we decided to adopt Dr. Murphy’s methodology and use her nutrient database as a starting point. Dr. Murphy generously offered the use of her Minilist and cross-reference file for this research project.
Our next steps included the evaluation of the vitamin D nutrient values in the Minilist and the cross-reference file created by Dr. Murphy. We first conducted an extensive comparison of the vitamin D nutrient values in the Minilist with vitamin D nutrient values listed in other sources, such as Bowes & Church (editions 1975, 1980, 1985, 1989, 1994), McCance and Widdowson's (edition 1991), the 1991 USDA Provisional Table on vitamin D content, and the nutrient database used by Dr. Jean Hankin at the Cancer Center of the University of Hawaii. It quickly became apparent that the vitamin D nutrient values listed for fish, one of the major dietary sources of naturally occurring vitamin D, vary by source (see table below), and that the comparison among sources is difficult as vitamin D nutrient values vary by type of fish (e.g., Atlantic herring vs. Pacific herring) or type of preparation (fresh vs. canned vs. smoked). Furthermore, none of the sources reviewed includes a complete list of all fish. Bowes & Church editions 1975 to 1994, for example, do not provide any vitamin D nutrient values for fish.

### Content of vitamin D (IU per 100 gram of food)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>raw herring</td>
<td>1,628 **</td>
<td>900</td>
<td>-</td>
<td>760 **</td>
</tr>
<tr>
<td>raw mackerel</td>
<td>360 ***</td>
<td>700</td>
<td>-</td>
<td>200 ****</td>
</tr>
<tr>
<td>smoked herring</td>
<td>120</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>canned sardines</td>
<td>272 #</td>
<td>300</td>
<td>500</td>
<td>300</td>
</tr>
<tr>
<td>raw Pacific salmon</td>
<td>500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>broiled salmon</td>
<td></td>
<td></td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>canned Chinook salmon</td>
<td>324</td>
<td></td>
<td>370</td>
<td></td>
</tr>
<tr>
<td>canned Pink salmon</td>
<td>624</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>halibut</td>
<td>600 ##</td>
<td>40 ###</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>oysters</td>
<td>320</td>
<td></td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

- * listed as raw Atlantic herring
- ** listed as raw Pacific herring
- *** listed as raw Atlantic mackerel
- **** listed as raw Pacific mackerel
- # listed as canned Atlantic sardines
- ## listed as canned Pacific sardines
- ### listed as Greenland halibut
- ### listed as Pacific halibut

We contacted Dr. Jean Pennington to inquire about the sources for the vitamin D values of certain foods included in her nutrient guide. We also inquired about the sources for the vitamin D values listed in Bowes and Church, as Dr. Pennington is the editor of editions 1985 to 1994. We learned that the Bowes & Church listing is not necessarily complete as it includes only foods for which the nutrient data were provided by the manufacturers. We also tried to identify the sources for the vitamin D values listed in the USDA provisional table and learned that the person who compiled the list is no longer with the agency. No information regarding the sources could be obtained.
We then turned our attention to vitamin D fortification practices in 1971-75. Foods which are fortified with vitamin D in the US include milk, cereals, margarine, and ovaltine. We contacted the Dairy Council and various manufacturers of cereals (e.g., Kellogg's, Quaker Oats, General Mills, Kraft General Foods), margarines, and milk flavorings, and inquired about the duration and amount of vitamin D fortification of specific brand name products reported in the NHANES I 24-hour dietary recall.

After compiling information on vitamin D from the various sources (published data, information provided by manufacturers) we reviewed the vitamin D nutrient values included in the Minilist and updated some values with those listed in the USDA provisional table and made some other modifications. We then carefully reviewed Dr. Murphy's cross-reference list, focusing our attention on the substitutions and recipes used for mixtures and made several modifications, which have been carefully documented.

We are currently finalizing this extensive review of the Minilist and cross-reference file. The task is about 95% complete. The addition of vitamin D nutrient values to the NHANES I nutrient database turned out to be considerably more difficult and time-consuming than anticipated. Our research efforts clearly demonstrate a lack of research on vitamin D nutrient values which greatly limits any research linking dietary vitamin D intake with specific health outcomes such as cancer, as well as the interpretation of previously published data on dietary vitamin D intake and specific health effects.

### 2.2.3. Identification of analytic cohort

NHANES I includes 8,596 women aged 25-74 years who completed the baseline interview and examinations. The following exclusions were made to establish the analytic cohort: (1) 814 women without a follow-up interview in 1982-84, 1986, or 1987 (self or proxy interview for deceased subjects); (2) 235 women who at baseline reported that they have had a prior malignancy (no information is available on year of diagnosis and type of malignancy); (3) 35 women who had some mention of breast cancer or multiple breast biopsies but for whom the date of breast cancer incidence could not be determined; (4) 15 women who reported a breast cancer which was determined to be prevalent. Thus, we excluded a total of 1,099 women from the analytic cohort. Additional exclusions (e.g., women without dietary data, women who were pregnant or breast-feeding during the 24-hour dietary recall, etc.) will further reduce the analytic cohort.

### 2.2.4. Identification of breast cancer cases

For the remaining eligible 7,497 women we carefully reviewed the interview and death certificate data for any mention of breast cancer. We identified a total of 190 women diagnosed with breast cancer during the follow-up period: 142 self-reports confirmed by hospital records, 33 self-reports without hospital record confirmation, 6 hospital record reports without self-report, and 9 death certificates listing breast cancer as the underlying cause (N=7) or a contributing cause (N=2) of death without confirmation by hospital records.
2.2.5. Estimation of person-years of follow-up

For women with breast cancer, the person-years of follow-up have been estimated from the date of the NHANES I interview and examination to the incidence date of breast cancer. The following data have been used as the breast cancer incidence date: the date of first hospital admission for breast cancer for self-reports confirmed by hospital records, the mid-point of the self-reported year of diagnosis (June 30) for self-reports without hospital record confirmation, and the date of death for the breast cancers confirmed by death certificates only.

For women without breast cancer, the person-years of follow-up have been estimated from the date of the NHANES I interview and examination to the date of last interview if alive or to the date of death if deceased.

2.3. Discussion

After completing the review of the Minilist and the cross-reference file and implementing various changes, the vitamin D content (per 100 grams of food) for each of the NHANES I foods codes will be added to the NHANES I nutrient database and the average dietary intake of vitamin D will be estimated for each individual in the analytic cohort. The database will then be ready to conduct the statistical analysis which is the focus for the second year of this project.
3. LITERATURE CITED


