D. Research Design and Methods

1. Design

This is a randomized, double-blind placebo-controlled study to test whether calcium (1200 mg/d) plus vitamin D (2000 IU/d) reduces the incidence of cancer in a population-based sample of women 60+ years of age. We will randomly sample the population of healthy, independent living women 60+ years in nine counties. We will randomly assign 2300 women to one of two groups: Group 1 will receive calcium (1200 mg/d) and vitamin D (2000 IU/d) and Group 2 will be given calcium and vitamin D placebos. Each participant will be followed for four years.

After all participants have completed four years of follow up, we will perform a nested-case control study to determine the association of serum 25OHD collected at randomization and at the end of year one of study with risk of cancer over four years. For each cancer case there will be two controls matched for age and race.

Table 4 on the next page indicates our goals for recruitment and retention of participants over the course of the study. Each project year is divided into quarters. The visit number is indicated with a “V”. The projected number of participants to be enrolled in a given quarter is indicated under V1, and that row shows the projected attrition for that group of participants over the remainder of the study. The total number of visits, as well as the subtotals by year and by quarter are provided at the bottom of the table.

Table 4

Timeline (recruitment over first 12 months)

Project years	Year 1				Year 2				Year 3				Year 4				Year 5			
Project months	1-3	4-6	7-9	10-12	13-15	16-18	19-21	22-24	25-27	28-30	31-33	34-36	37-39	40-42	43-45	46-48	49-51	52-54	55-57	58-60
Visit number	V1		V2		V3		V4		V5		V6		V7		V8		V9			
Subjects	383		377		370		364		357		351		345		339		333			
Visit number		V1		V2		V3		V4		V5		V6		V7		V8		V9		
Subjects		767		753		740		728		715		703		690		679		667		
Visit number			V1		V2		V3		V4		V5		V6		V7		V8		V9	
Subjects			767		753		740		728		715		703		690		679		667	
Visit number				V1		V2		V3		V4		V5		V6		V7		V8		V9
Subjects				383		377		370		364		357		351		345		339		333
Visits by quarter	383	767	1143	1137	1124	1117	1104	1098	1085	1079	1066	1060	1048	1042	1030	1024	1012	1006	667	333
Visits by year				3430				4443				4290				4143				3018

Total entered	2300
Total finished	2000
Total lost	300
Lost as % of entered	13.0%

GRAND TOTAL VISITS (all years) 19324

2. Sample

a. Participants

Study participants will include a random sample of women at least 60 years of age who are in good general health, are at least 4 years past last menses, and are living independently in a nine-county rural area of Nebraska. They must be able to travel to the Fremont Area Medical Center (FAMC) for study visits. Exclusion criteria include: 1. history of cancer except for superficial basal or squamous cell carcinoma of the skin or other malignancies treated curatively more than 10 years ago; 2. history of chronic kidney disease; 3. participation on the previous population-based calcium and vitamin D study.

b. Sample Size and Power

Primary Hypothesis

Our primary hypothesis is: Increasing serum 25OHD from prevailing levels will significantly decrease incidence of all types of cancer combined in a population sample of healthy postmenopausal women. The State of Nebraska age-specific cancer incidence rates (excluding non-melanoma skin cancers) for women aged 60-69 is 1.2%/yr; for women aged 70-79, 1.7%/yr, and for women 70+, 1.8%/yr.[97] The overall annual cancer incidence rate in our previous (pilot) study cohort was 2%/y in the control group, 0.5%/y in the calcium/vitamin D treatment group, and 1.2%/y overall. These observed values are consistent with population level rates, indicating that cancer rates in our pilot study were about what should have been expected. In our pilot study, with a substantially smaller sample, we had a *post hoc* sample power of about 66% to find an unadjusted risk ratio of 0.4 (the actually observed value).

Since the effect of vitamin D in the first year would be expected to be slight, power was estimated using several scenarios involving different sample sizes, effect sizes, and incidence rates. In doing so we have assumed: 1) that the hazard rate ratio would change from 1.0 during the first six months of observation, to the specified effect size by the third 6-month period, with the second 6-month period at an intermediate rate (as we had observed in the pilot study); and 2) a rate of loss of subjects from study of 1.5%/period (also about what occurred in our pilot study). Calculations were performed using the Survival Analysis module of SamplePower™ Release 2.0 (SPSS, Chicago). [98-101] This program uses the Arcsin method described at http://www.uoregon.edu/~robinh/arcsin.txt.

For base cancer rates of 2.0, 1.5, and 1.0%/y, and an enrolled sample size of 2,300, power is as follows:

Table *5*	**Base Cancer Rate (y^{-1})**		
Effect Size (RR)	**2%**	**1.5%**	**1.0%**
0.6	62%	50%	36%
0.5	82%	71%	54%
0.4	95%	87%	72%
0.25	100%	98%	92%

Judging from official Nebraska statistics, the rate is not likely to be as low as 1.0%, and would more likely be 1.5% or greater, depending upon the final age distribution of the sample.

As noted, the unadjusted effect size in our pilot data was 0.4. If approximately the same reduction occurs again, as the Table *5* shows, we would have a power of 95% at a base cancer rate of 2%/y, 87% at a base cancer rate of 1.5%/y, and 72% at a base rate as low as 1.0%/y. By contrast, the adjusted effect size (RR) in our pilot data was ~0.25, and if the approximately the same reduction obtains in this replication study, even at a base rate as low as 1.0%/y, we would have ample power using two samples of 1150 each. Hence, balancing the various tradeoffs, such a sample size should provide sufficient power to find the reduced cancer incidence noted in our pilot data. Based on 13% attrition in our pilot study, we will enroll 2300 participants to end with a sample of at least 1000 per treatment group.

Secondary Hypothesis:

Our first secondary hypothesis is: Increasing serum 25OHD from prevailing levels will significantly decrease incidence of specific cancers: breast, lung, colon and myeloma, leukemia, lymphoma. We analyzed subgroups with specific types of cancer in our original cohort limiting the analysis to those free of cancer after one year of study. In comparison to placebo, the relative risk (RR) of developing breast cancer for the Ca+D group was 0.22 (C.I.: 0.043 to 1.079, P = 0.062) Neither of the

Table 6					
Cancer Site	***P***	***RR***	***CI***		***Power α=0.10***
			Lower	***Upper***	
Breast	***0.062***	***0.216***	***0.043***	***1.079***	***89%***
Lung	***0.188***	***0.218***	***0.023***	***2.107***	***81%***
Lymphoma, Leukemia, Myeloma	***0.367***	***0.437***	***0.073***	***2.634***	***33%***

other subgroups, lung or lymphoma, leukemia, myeloma, demonstrated a statistically significant reduction of risk with vitamin D. Lung cancer came the closest to having a significant effect, RR = 0.22, C.I.: 0.023 to 2.107, P = 0.188. A relative risk after the first year cannot be computed for colon because there one only one case. In Table 6 we show our estimates of power to find significant differences in cancer risk between the two treatment groups in the proposed study with a sample size of 1000 per group.

c. Recruitment of Participants

Midwest Survey and Research (MSR) will conduct the major portion of the recruitment for the proposed study. MSR is a nationally recognized market research firm that successfully completed our recruitment of 1180 women in our previous population-based study. MSR has developed the plan (See Appendix) for randomly sampling the population of independently living postmenopausal women in a nine-county rural area of Nebraska. The MSR Group will pull every listed household in the 9 county sample area. There are an estimated 51,000 households in the sample area. To save cost, and because the sample area has a very high listed phone number ratio, random digit dialing (RDD) will not be utilized. After every listed household is pulled, the list will be completely randomized.

Once the list is randomized, MSR will use replicates. A replicate is a sample within a sample. It is randomly drawn from the original list. Replicates are used to control fielding to ensure that the calling effort has been consistent across the entire geography of samples. MSR will pull replicates in increments of 1000. Each one of those 1000 numbers will be attempted up to five times before being discarded. All 1000 numbers must be used up before the next replicate is selected. Any number that is consistently a no-answer or an answering machine, will be isolated and attempted several months later. MSR will continue to pull replicates in increments of 1000 from the original base of all listed households until the study has been completed. RDD will not be done because comparison of listed numbers with census data indicates that more the 95% of the households have listed numbers. *Even though many in this population have cell phones, they maintain their landlines.* Thus, our population will be almost completely included in the random sample. Furthermore, random digit dialing would be very expensive since estimates are that MSR would make 30 calls to get a hold of one real household number. Many of these rural prefixes include only about 300 numbers compared to Omaha in which one prefix includes 5000-6000 numbers. Thus many of the generated numbers would be non-existent. Thus RDD would incur much cost for little benefit.

All telephone interviewing will be conducted by professionally trained MSR interviewers. An initial telephone screen will eliminate most of the candidates who fail to meet the entry criteria. MSR will track the number of persons who fail the entry criteria, the number who meet criteria but chose not to participate, and the reasons given for refusal. Those who pass this screen and agree to participate will be will be contacted by the project staff to set up an appointment for further screening, informed consent and baseline measurements.

In order to develop community awareness of our study and allay any potential suspicion about the phone call from MSR, we will send letters to physicians in the nine county study area informing them of the study. Thus, they will be alerted lest their patients ask their advice about joining the study. We will also work with Creighton and the FAMC, the study site, to develop a press release in the study geographical area prior to commencement of study recruitment. Our study staff will call back all interested MSR-screened parties within 72 hours of obtaining their names to give them a brief explanation of the study and schedule a screening visit.

3. Study Site

We will set up a study site at the Fremont Area Medical Center (FAMC) in Fremont NE, which is in the central part of the nine-county rural area we are sampling. We find that rural women in NE are reluctant to travel to Omaha to participate in research studies. We conducted our pilot study at FAMC and had exceptionally good adherence. Participants rarely missed visits and overall attrition was very low. We have approval from the CEO of FAMC to conduct the study there if funding is awarded. (See letter of support.)

4. Assessments

a. Cancer Ascertainment

During each semiannual study visit, participants will be asked whether they were diagnosed with cancer since the last study visit and the name of the physician who provided care for her. The study participant will then sign a release of information form that allows the physician/hospital who provided the treatment for the cancer to provide us with the medical records report of the cancer as well as the specific ICD-9 cancer code. *We will arrange for central adjudication of the pathology. On the HIPPA form that subjects sign along with consent at the beginning of the study, we will ask the subjects to give approval to send any cancer pathology*

slides to the adjudicating pathologist. Our method is intended to determine the number of true positives that agree with the study participants' reporting of cancers. We are also able to determine the number of false positives.

We will also determine the number of false negatives, i.e., those who were diagnosed with cancer but did not report it during an annual visit. To determine this, we will send a list of each physician's study participants to him/her and ask each physician to verify the fact that each of his/her patients was **not** diagnosed with cancer during the period of study. We will obtain a release of information from each study participant to obtain this information from her personal physician. We will randomly sample 10 percent of those not reporting cancer diagnoses and query their physicians regarding any cancer diagnosis/treatment during the study period.

In addition, and if necessary, to the above-mentioned procedures for securing cancer incidence data, we will use a standard epidemiologic procedure to obtain otherwise missing information, i.e., to contact designated "next of kin". Relatives' names are obtained at study initiation, and study participants are told that next of kin will be contacted only if the study participant is not available, at the end of the study. We will use this procedure for those study subjects who may not complete the study. We contact the next-of-kin to verify cancer occurrence and obtain permission to contact the personal physician. We would request from the physician the medical report as well as the specific ICD-9 code of the cancer.

b. Laboratory Tests

25OHD

We will analyze serum 25OHD in each participant annually as it is the best biomarker of vitamin D status and a predictor variable we would expect to incorporate into our analyses at end of study (as we did for the pilot study). We also will use it to assure that the treatment group is achieving an increased vitamin D level and to assess the vitamin D status of the placebo group. We will use values obtained at baseline and at the end of year one of study in a nested case control study in which we determine the association of serum 25OHD with incidence of cancer. 25OHD will be measured by radioimmunoassay after extraction using the IDS kit (Nichols/Quest Diagnostics, San Clemente, CA). Our laboratory participates in the DEQAS quality assurance system for 25OHD assays, and over the course of the study our findings on test samples were regularly close to the international mean. Our laboratory technologist will be the only person unblinded to the serum 25OHD levels. Dr. Recker will be unblinded only to values out of the clinical reference range, which he will evaluate. In our previous study, out-of-range values were rare.

1,25OHD2

Since analyzing serum 1,25OHD$_2$ may not yield helpful information and is expensive , we will collect samples at each annual visit and store them. If our study findings show that serum 1,25OHD$_2$ would be of interest, we will seek funding to pay for the analyses. The rationale for saying that it may not yield helpful information is as follows: The active form of vitamin D is 1,25OHD$_2$. However, it is hypothesized that many tissues and cells, including those in the breast, colon, and prostate that are common cancer sites, can make their own 1,25OHD$_2$ (autocrine function). This 1,25OHD$_2$ is capable of regulating cellular processes that help to maintain cellular growth and prevent malignancy.[29;102] Amounts of this locally produced 1,25OHD$_2$ are not measurable in the serum. Serum levels of 1,25OHD$_2$ are not sufficient to provide intracellular concentrations needed for optimal autocrine function in the specific epithelial tissues.

Safety lab

We will analyze serum calcium and creatinine annually to monitor for hypercalcemia and changes in renal function.

Blood Storage

We will keep duplicate blood samples for all subjects to be used as blinded quality control samples for the vitamin D assays or other assays. We will also collect and store 30 cc of blood from every participant to test for genetic markers should the intervention be found effective in decreasing the incidence of cancer. Blood will be stored in a -70 freezer.

c. Medical and Social history

At baseline we will obtain a thorough medical and social history. We will ascertain the time interval since the last mammogram and most recent screening exams (fecal occult blood, sigmoidoscopy, barium enema, colonoscopy, or virtual colonoscopy) for colon and rectal cancer, Pap smear, endometrial biopsy, and full skin examination for cancer. We will ascertain premalignant lesions that have been identified and removed. We will include that information in stratifications and multivariate analyses, allowing us to adjust for time since last screening test. If needed, we can examine the effect in women who have been screened recently. We will also ascertain a full history of use of replacement hormones by each woman. The history will also provide other

covariates for the analysis, such as family history of cancer, personal history of illness, medication use, etc. We will update this information at semiannual study visits. We will thoroughly assess medication and supplement use in the subjects at baseline and during the study. We will query the subjects at 6-month visits and ascertain start and stop dates as well as dosage. We record the name of the medication/supplement and code it using the American Hospital Formulary Service pharmacologic-therapeutic classification system.

As we did in our previous study, we will ascertain all new diagnoses during the study and collect related clinical information. We will designate as specific secondary chronic disease endpoints new diagnosis of the following: hypertension [1], cardiovascular disease [1], osteoarthritis [2], colonic adenomas[3] and diabetes[4;5]. We will obtain medical records to confirm new diagnoses. We will also ask subjects about acute viral symptoms, such as upper respiratory diseases and influenza.[6] Dr. Recker will serve as our adjudicator to designate those diagnoses that will be acceptable as incident chronic diseases or acute viral diseases. He will be blinded to treatment assignment.

d. Height and weight

At baseline and annually we will measure height and weight. These variables, plus calculated body mass index (BMI), will be used as covariates in the analysis. Weight change will also be monitored.

e. Dietary intake (FFQ).

We will administer the Block 2005 Food Frequency Questionnaire (Block Dietary Data Systems, Berkeley CA). The FFQ will enable us to estimate intake of all nutrients in this cohort so that we can use nutrient estimates as covariates in the analyses. This full-length (approximately 110 food item) questionnaire was designed to estimate usual and customary intake of a wide array of nutrients and food groups.(See Appendix.) It takes 30-40 minutes to complete and is intended for either self- or interviewer-administration. Both the food list for this questionnaire and nutrient database for its analysis were developed from more recent data than used for the Block '98 FFQ. The food list was developed from NHANES 1999-2002 dietary recall data; the nutrient database was developed from the USDA Food and Nutrient Database for Dietary Studies (FNDDS), version 1.0. A series of "adjustment" questions provide greater accuracy in assessing fat and carbohydrate intake. Individual portion size is asked for each food, and pictures are provided to enhance accuracy of quantification. We will use the booklet rather than the on-line version due to the booklet being much less expensive. Block Dietary Data Systems (Berkeley CA) will analyze the data and provide us with reports.

f. Physical activity

We will measure physical activity at baseline and at the final visit to use as a covariate in the analysis, since it has been shown to have an inverse association with cancer. [103;104] We will use the Short Last 7 Days Self-Administered version of the International Physical Activity Questionnaire (IPAQ).[105] (See Appendix.) The tool was developed to obtain internationally comparable data health-related physical activity in adults. It has undergone extensive reliability and validity testing. Test-test reliability indicated a Spearman correlation coefficient of 0.88 in the U.S.[105] Typical correlations for criterion validity had a median value of 0. 30. The form includes seven questions and is easy to administer. Thus, it is appropriate for the large number of subjects in the proposed study.

g. Adherence with supplementation/placebo

We will determine adherence by weighing each bottle of tablets before distributing the bottle and after it is returned by the participant at each study visit. We will initially determine the weight of each pill: calcium, vitamin D_3, placebo calcium and placebo D_3. To do this, we weigh the empty bottle and then the full bottle. Then we subtract the weight of the empty bottle from the weight of the full bottle and divide the difference by the number of tablets in the bottle. We have used this method in several studies and find it to be more accurate and more time efficient compared to pill counts for large studies.

We provide our subjects with much support to promote adherence. We encourage them to tell us difficulties they encounter with the supplements so that we can advise them about how to overcome any obstacles. We consistently have excellent adherence, as noted earlier for our first study. We schedule study visits every 6 months to promote adherence.

h. Falls

We will also assess for falls with the fall assessment tool that we used in our previous study. [7] We will interview participants at each visit about falls experienced over the last six months. Participants will be asked to keep a falls diary noting the date of the fall. They will also be asked to note whether the fall was indoors or outdoors and whether it resulted from a slip, trip, an external force (such as a dog), an intrinsic reason (such as dizziness) or no identifiable reason.

5. Intervention

We will randomly assign 2300 women to one of two groups: Group 1 will receive calcium (1200 mg/d) and vitamin D_3 (2000 IU/d), and Group 2 will be given calcium and vitamin D placebos. Each participant will be followed for four years. Calcium will be in the form of calcium carbonate 600 mg caplets or identical placebo, and each subject will be given two caplets/d, one with breakfast and the other with the evening meal. *Recommended levels of calcium intake were revisited by the RDI Committee and set as 1,200 g/day[16], the level of supplementation that we are including in the proposed study.* Vitamin D_3 will be made by Tishcon Corp (Westbury, NY) supplied such that each capsule will have at least 2000 IU. Each subject will receive one capsule/d or identical placebo. We will use each batch of capsules within one year. We will analyze a sample of each lot of vitamin D_3 when we receive it and at the end of each year to assure potency. We used this product in our pilot study and found that capsules retained potency for more than one year. Vitamin D and calcium placebos will be identical to the active supplements and will contain the same excipients as the active.

Participants will be informed that they may be assigned to placebo calcium and vitamin D. Because of the well-established need for calcium for optimal skeletal health, participants will be told that they may take their own calcium supplementation in addition to the study pills. However, to avoid risks of excessive calcium intake, participants will be advised that they should limit their outside calcium supplementation to 500-600 mg/day. This combined with their dietary intake of calcium will meet the current recommendations for health. Participants will also be allowed to take their own vitamin D, but we will ask them to limit that to no more than 400 IU/day if they are < 70 years of age and to more than 600 IU/day if they are ≥ 70, which are the currently recommended intake levels.

a. Randomization

We will need 2000 study participants to achieve more than 80% power and, since we expect 13% drop-out rate through the whole study (based on our pilot study), the final total sample size comes to 2300. The planned 2300 study participants will be randomized at their recruitment to one of the two treatment groups (Placebo or Vitamin D/Calcium). We will use block randomization, with block of length 2. That is, patients will be divided into pairs, according to their arrival order, and will be randomized within their respective pair. The study is double-blinded, i.e. neither the study participants nor the PI and her team will know the exact treatment, except for the study statistician.

b. Procedure

The population of healthy independently living postmenopausal women 60+ years of age living in a nine-county area of NE will be sampled by MSR. Interviewers at MSR will contact women from the sample and conduct an initial screening over the telephone. The names and contact information of women who pass the initial screen will be forwarded to the Creighton study staff, who will then conduct a second telephone screen. Eligible candidates will be scheduled for a baseline visit. The study will be explained and consent will be obtained. At the baseline visit, the following procedures will be performed: assessment of medical and social history; measurement of height and weight; completion of dietary and physical activity questionnaires; and phlebotomy. The subjects will be enrolled, randomized to treatment group, and provided calcium and vitamin D_3 supplements/placebo. Study visits will be scheduled semi-annually. At each visit, the following will be assessed: recent medical and social history, including adverse events; ascertainment of cancer diagnoses; and adherence to supplement/placebo. Study supplements will be replaced at each visit. Annually, phlebotomy will also be done to obtain samples for 25OHD, 1,25D, calcium and creatinine. Study visits will take 30 to 45 minutes.

Adverse will be evaluated by Drs. Lappe and Recker, and serious adverse events and unexpected events will be reported to the Creighton IRB and the Data Safety and Monitoring Board. Participants will be referred for follow up if appropriate. An intense effort will be made to maintain every participant on study until the project is completed. We will attempt to maintain contact with any participant who does drop from study and to ascertain diagnosis of cancer at the time they would have completed the study.

6. Data Analyses

The statistical analysis will focus on testing our

Primary Hypothesis: Increasing serum 25OHD from prevailing (control) levels will result in a statistically significant decrease in the incidence of diagnosis of all types of cancer combined in a population sample of healthy postmenopausal women by the end of the four-year follow-up period.

Since we expect some dropouts in the course of the study, an appropriate main endpoint of interest will be of time-to-event type, where the event is cancer diagnosis for the first time and the time to diagnosis is measured from each subject's own date of entry. All subjects who do not experience the event will be right censored at either the end of the study (if they were followed up through the whole study) or at the time point when they were lost to follow up (e.g. because of death from other causes, moving out of the area, unable to continue on study). Censored observations cannot be ignored in the analyses. The method of analysis must take the censoring into account and correctly use the censored observations as well as the uncensored observations. Therefore, our main approach to statistical modeling of the data will be survival analysis. [106] Its purpose will be to model the underlying distribution of time to diagnosis and to assess the dependence of the failure time variable on the independent variables, i.e. the treatment and other covariates. (We will use the expressions time-to-diagnosis and survival time interchangeably.)

While the effect of treatment (increase in serum 25OHD levels vs. otherwise) will be of principal interest, covariates that are related to carcinogenesis and cancer diagnosis (e.g., race, dietary intake of vitamin D, calcium, retinol, soy products, etc; age; smoking; BMI, physical activity, hormone replacement therapy, parity, family history of cancer) will be also included and measured at study outset. Cox regression [107] is a well-established general approach whose extensions will serve our purposes very well. The regression coefficient in Cox regression is interpreted as the change of the log hazard ratio resulting in the increase of one unit in the covariate. For each particular subject, some covariates will stay fixed through the study while others may change through it (e.g. switching treatment groups for some patients who are on and off medication) which will require the use of Cox regression with time-dependent covariates.

In addition to Cox regression, which is of the relative risk type, we will use the Accelerated Failure Time (AFT) model, [108] which postulates direct relationship between survival time and covariates. The latter has easier interpretation, and the results from it will be used to cross-check the ones from Cox regression. The AFT model is a parametric model, and the distribution of its error term of the model needs to be determined for appropriate interpretation of the coefficients of the regression part of the AFT model. Using goodness-of-fit procedures and graphical methods to evaluate the model fit, the best distribution will be selected from the standard ones for PROC LIFEREG in the statistical package SAS® 9.1 (see comments below). Once the parameters of the distribution are known, hypothesis testing will proceed as implemented in PROC LIFEREG for the AFT model. All P values will be based on two-sided tests and considered significant if less than 0.05. The statistical package SAS® 9.1 (SAS Inst, Inc, Cary, NC) will be used in all analyses. There are three SAS procedures that implement both basic and sophisticated models for analyzing survival data: LIFEREG, LIFETEST and PHREG. PROC LIFEREG is a parametric regression procedure for modeling the distribution of survival time with a set of covariates (some of which may be categorical), and will be used to implement the AFT model. PROC LIFETEST is a nonparametric procedure for estimating the survivor function (Kaplan-Meier estimator [109] or the life-table method), comparing the underlying survival curves of two or more groups, and testing the association of survival time with other variables. PROC PHREG is a semi-parametric procedure that will be used to fit the Cox regression model, with or without varying covariates.

Cigarette smoking status will be categorized as never, former, or current smokers and updated annually. A more detailed categorization will be based on the number of cigarettes smoked per day among current smokers (0, 1-14, 15-34 and 35+ cigarettes per day), as has been done in other cohort studies[27]. Pack years will also be calculated.

In addition to the survival analysis approach, we will perform logistic regression analysis of the event cancer diagnosis (Yes vs No) on treatment, while adjusting for covariates, similar to the survival analysis. To this end, we will compute person-time of follow-up for each subject from date of study entry to the date of cancer diagnosis, death from any cause, or the end of follow-up, whichever comes first. Incidence rates of cancer will be obtained by dividing the number of incident cases by the number of person-years in each category of exposure. That is, our research unit will be person-year rather than person only. In particular, we will obtain maximum likelihood estimate of the relative risk (RR) for each upper exposure category by dividing the incidence rate in that category by the rate in the lowest category, through a generalized linear model as implemented in the SAS procedure GENMOD.

We do not plan an interim analysis, but we have provision for it in the event the DSMB should want it. (See below.)

Interim Analysis. Stopping the study prematurely has a serious potential to bias the estimates of the statistical models parameters (Risk Ratio, etc) and influence their credibility. Interim analysis (of treatment comparison for efficacy and safety) may result in breaking the blind and premature stoppage of the study. At

the very least, it will affect the probability of Type I error and may decrease the statistical power of the study. Nevertheless, we plan to conduct interim analysis at the middle of the study (i.e. at the end of the 2nd year of treatment for all subjects), and the Data Safety and Monitoring Board (DSMB) will be informed accordingly. The interim analysis will be confidentially performed by the study statistician, Dr Haynatzki, and will be reported at a closed session of the DSMB. (No unplanned interim analyses will be conducted.) The early-stopping rule/algorithm based on calculating an alpha-spending function is provided below. We have applied the widely accepted approach that makes use of an alpha-spending function $\alpha'(t)$ that is defined for t=0, 0.5, 1.0 [110]. By definition, $\alpha'(0)=0$ and $\alpha'(1)=0.05$, and we further specify $\alpha'(0.5)=0.0025$. That is, when the interim analysis is conducted at time t=0.5, we will use the above alpha-spending function $\alpha'(t)$ to compare Control vs. Treatment. If this comparison results in a significant difference at level α', we will report this to the DSMB.

Nested case control study. Incident cancer cases that are newly identified during years 2, 3 and 4 will be included in this study. We expect a total of 75 incident cancer cases, according to percentages from our pilot study. These 25 cases per year are expected to include some 5 (i.e. 0.5%) of the 1000 effective vitamin D treated subjects as well as some 20 (i.e. 2%) of the 1000 effective controls. The total of approximately 75 new incident cancer cases will be matched to normal controls in a 1:2 matching, according to age and race. Values of 25OHD will be seasonally adjusted as we did in a previous analysis. [111] The matching on age will be done within 2-year intervals (i.e. within 1 year older or 1 year younger than the respective case). There are several possible ways to analyze the so-obtained matched pairs data, in order to test for association between serum 25OHD levels and cancer incidence. The simplest approach is to first average the serum 25OHD level (at year 1) of the 2 controls for each case, and find the difference between the so-obtained average and the value for the case. Next, perform a two-sided *t*-test on the sample of 75 differences with null hypothesis that the parental population has mean zero (0). In case of acceptance of the null hypothesis (at level of significance 0.05) we will conclude that the data at hand do not support association between serum 25OHD level and cancer incidence, whereas rejection of the null hypothesis will imply association. Another, more comprehensive approach would be to develop a linear model (not for the difference but) for the raw serum 25OHD level as response variable, which model tests the effect of Cancer status (Yes vs No) while adjusting for covariates on which we did not match (e.g. smoking; BMI, physical activity, hormone replacement therapy, parity, family history of cancer), and accounting for the fact that the data are matched pairs. A still more comprehensive analysis, instead of the raw serum 25OHD at year one, is to consider the difference between year one and baseline, and build a model similar to the previous one.

***Secondary Hypotheses**:*

1. Increasing serum 25OHD from prevailing levels will significantly decrease incidence of specific cancers: breast, lung, colon and myeloma, leukemia, lymphoma. Based on our previous study, we expect the most commonly occurring cancers to be breast, lung, colon and lymphoma, leukemia, myeloma. Therefore, we will have each of these as secondary endpoints. We will test differences in frequency of breast, lung, and colon cancer and lymphoma, leukemia, myeloma between the two treatment groups by Fisher Exact tests if possible, otherwise chi-square with Yates's correction for continuity (small numbers). We will use proportional hazards regression for these cancers and for any other frequently occurring cancers where sample size allows. Also, we will divide both baseline and one-year values of serum 25(OH)D of the entire cohort into quintiles and determine whether there are statistically significant differences between the upper and lower quintiles in cancer risk of the entire cohort for the specific cancer endpoints. We will also test for trend w.r.t. quintiles, which will be more statistically powerful. Even if we do not find statistically significant differences between the specific cancer subgroups, the analyses will still be of interest for generating new hypotheses and planning new studies.

2. Increasing serum 25OHD from prevailing levels will significantly decrease incidence of other disorders, specifically hypertension, cardiovascular disease, osteoarthritis, colonic adenomas and diabetes, upper respiratory infections and falls. We will analyze for these disease endpoints similarly to the main endpoint (all type cancer diagnosis). Although the study is not powered to find statistically significant differences between the treatment groups for these non-cancer endpoints, we may be able to detect an effect of vitamin D supplementation. We will use Fisher Exact tests if possible, otherwise chi-square with Yates's correction for continuity (small numbers), to determine differences between the two treatment groups in frequency of the specific diseases. We will use proportional hazards regression for hypertension, cardiovascular disease, osteoarthritis, colonic adenomas, diabetes, falls and acute viral disorders. Also, we will divide both baseline and one-year values of serum 25(OH)D of the entire cohort into quintiles and determine whether there are statistically significant differences between the upper and lower quintiles in diagnosis of the designated

diseases of the entire cohort. We will also test for trend w.r.t. quintiles, which will be more statistically powerful. If we find associations or significant trends between supplementation and/or serum 25(OH)D levels and other diseases, future studies to target a specific disorder can be developed.

7. Data Management

The study documents (hard copy) will be completed by the study nurses. All participants' case report forms (CRF's) will be stored in the project nurses' office. A Microsoft Access database will be developed with forms that match the CRF's.

All handling of electronic data will be by forms. The study site personnel do not have access to the underlying tables or to the database window. All data entered into the electronic data base, aside from names and personal identifiers, will have validation rules. For instance, [date of birth] must fall into a permissible range, and each such field will have a validation message to prompt the user when an impermissible value is entered. Entries for categorical variables, like [recruitment status] or [medications], will be selected from pre-established lists stored in tables rather than just typed in.

All non-null data will be audit-trailed. Once a value has been entered or selected for a field, a change in that value will prompt a preemptory demand for explanation. This will result in storing in an Audit table a new row of data containing the table name, the field name, the user id, the date and time, and the reason. Periodically, auditing queries will be run again in the Audit table to see if anything unusual is happening with respect to particular data fields, sites, or participants.

We have up-to-date anti-virus and firewall protection for the project staff members' workstations. Physically, the workstations are in locked offices, and each is password protected. The workstations are connected in a local area network (LAN) through a router under our control. The database is stored on one of the workstations. This workstation has daily incremental and weekly full backups to a tape drive. For off-site backup, the database is transmitted encrypted at the end of each day to a file server complex in the Creighton computer center complex, where it has the security projection of that center. There is also a laboratory database for the assays of the serum samples. This is maintained through a secure workstation in our biochemistry lab, and it resides in the same file server complex, which is a mile distant from the biochemistry lab.

To assure that the data are transmitted reliably from paper into electronic form, Dr. Haynatzki will supervise quarterly checks of the data during the recruitment period. At each check, 10% of the biweekly dataset will be randomly sampled by him and checked for errors by a data entry person. All errors that are detected in the sample will be noted. Based on the sample, a 95% confidence interval (C.I.) will be built for the proportion of errors in the quarterly dataset. If the upper bound of the C.I. exceeds 0.05%, the whole quarterly dataset will be rejected and re-entered completely; otherwise, accepted. In any case, all detected errors will be corrected by the data entry person.

8. Potential Problems for the Study

a. The sample of 2300 is a large number to recruit in one year.

We recruited 1180 women in one year in our pilot study. We are working with MSR, the marketing company with whom we worked in the pilot study so we are confident they can provide the candidates in a timely manner. The most recent census data indicate that there are more than 26,000 women 60+ years in the nine-county area that we are sampling. Our experience is that older women in this community are very interested in participating in studies, even studies requiring more involvement than this one.

b. We are not measuring biomarkers other than 25OHD.

We are collecting and storing blood and white blood cells to test for genetic markers should the intervention be found effective in decreasing the incidence of cancer. Blood will be stored in a -70 freezer. Markers selected will depend upon the state of the science at the time of analysis.

c. The sample is limited to post-menopausal women

Findings from our pilot study, which included only post-menopausal women, need to be confirmed with a prospective study designed with incidence of cancer as the outcome. Positive findings from the proposed study will provide support for future studies in men and younger women.

d. We are not performing any cancer screening tests to detect early cancers.

This would increase the cost prohibitively. In any event, we are interested in all cancers as the outcome, and there are no accepted mass screening tests with adequate sensitivity and specificity for early lung and

bronchus, pancreatic, ovarian, esophageal, or gastric cancer, leukemia, lymphomas, or multiple myeloma, among others. We will ascertain the time interval since the last mammogram and most recent screening exams (fecal occult blood, sigmoidoscopy, barium enema, colonoscopy, or virtual colonoscopy) for colon and rectal cancer) Pap smear, and endometrial biopsy, and full skin examination for cancer, will include that info in stratifications and multivariate analyses, allowing us to adjust for time since last screening test. If needed, we can examine the effect in women who have been screened recently. We will also ascertain a full history of use of replacement hormones by each woman.

e. We are allowing subjects to take their own calcium and vitamin D supplements.

Because of the high incidence of osteoporosis in this population and because some subjects will be given placebos, we believe it would be unethical to advise women to avoid self- supplementation. Participants will be informed that they may be assigned to placebo calcium and vitamin D. Because of the well-established need for calcium for optimal skeletal health, participants will be told that they may take their own calcium supplementation in addition to the study pills. However, to avoid risks of excessive calcium intake, participants will be advised that they should limit their outside calcium supplementation to 500-600 mg/day. This combined with their dietary intake of calcium will meet the current recommendations for health. Participants will also be allowed to take their own vitamin D, but we will ask them to limit that to no more than 400 IU/day if they are < 70 years of age and to more than 600 IU/day if they are ≥ 70, which are the currently recommended intake levels. We did allow subjects in our previous study to take their own supplements. We monitored self-supplementation, quantified it and factored it into the analysis. The median and interquartile range for calcium supplementation was 375 mg/d (0-762) and for vitamin D supplementation it was 200 IU/d (0-400). In those taking vitamin D, the mean 25OHD at baseline was only 13.4 nmol/L higher than in those not receiving vitamin D (75.7 nmol/L vs 62.3 nmol/L respectively); while treatment produced a rise of about 25 nmol/L, nearly twice as great. In the proposed study, with twice the dose, the increment will be correspondingly higher. Thus it is a virtual certainty that we will have a substantial difference in achieved D levels between the contrast groups. With nutrients such as D, it is never a matter of comparing something with nothing (as in drug studies), but more with less. Vitamin D self-supplementation may increase because of our findings. Many supplements contain vitamin D_2 which is less active than vitamin D_3. We will be able to track and quantify vitamin D usage by interview every 6 months and by analysis of serum 25OHD levels annually. In a secondary analysis, we will look at the association of serum 25OHD and cancer incidence. It is widely acknowledged that on average humans are not adherent to medication or supplement regimens. In fact, in the WHI clinical trial of vitamin D supplementation, adherence was only about 50%. Thus, we believe that allowing self-supplementation will not adversely affect our study, and we anticipate no effect on power.

f. One of the reviewers has stated that " this is really an ambitious phase III trial. However, whether it will yield additional critical information on chemoprevention on calcium and vitamin D is questionable".

The study may be considered ambitious in the context of chemotherapy Phase III studies, but it is not large in the context of primary prevention trials. For example, the Women's Health Initiative (WHI), which was a primary prevention trial, enrolled 29,000 women in its calcium-vitamin D arm; this trial proposes enrolling only 2300 women, or less than 10% of the women enrolled in the WHI.

_This study provides additional critical information since it uses a higher dose of vitamin D3 (2000 IU) than has been used previously as part of an intervention in a study with cancer as a designated outcome. This is new and critical because it will help to quantify the dose-response relationship of vitamin D with incidence of cancer. Previous studies have used only 1100 IU/day[17] or a very low dose of vitamin D (400 IU).[18]

_The timing of this study is crucial. As yet, no voluntary or professional medical organization in the US recommends intake of vitamin D3 in the 2000 IU dose range for disease prevention, so the study is feasible now. If the study is delayed, this could change, and the study would become far harder to perform. The result would be the loss of a great opportunity for confirmation of the effect of 2000 IU vitamin D and calcium in a human clinical trial. This would be a loss to the contribution of science to the vitamin D-calcium-cancer issue.

_The investigators are experienced in doing clinical trials, and have the requisite skills and knowledge for successful completion of this study.

_The study design is optimal for answering the critical question that is being posed. Specifically:

1) The proposed study will have high statistical power for detection of an effect of the intervention on all cancers (87%-98%), assuming the anticipated base cancer incidence rate of 1.5% per year, and effect size (RR=0.25) is as previously observed. Power is also adequate (87%) for a detection of a much weaker than expected effect size (RR=0.4), based on an estimated base rate cancer incidence rate of 1.5% per year. Since cancer incidence is exponentially associated with age, and the new study includes a slightly older age

distribution (>60 vs. > 55 years in the previous study by our group), this makes it highly likely that the base cancer rate is at least 1.5%.
2) The latitude of Omaha is 41.4 degrees, and its climate is consistent with a sufficient duration of the continental winter climate to allow a drop in summer 25(OH)D, which has a half-life of only 3 weeks, making the effect of the intervention dose of vitamin D more detectable than in an area with a shorter winter.
3) The study uses vitamin D3, which was used in the previous study by our group and is consistent with most observational studies that have shown an effect of vitamin D on risk of cancer, where it is possible to at least approximately distinguish vitamin D3 from vitamin D2.
4) The vitamin D content of each lot of vitamin D tablets will be confirmed when received and at the end of each year. The vitamin D tablet are from the same manufacturer (Tishcon) that provided the tablets for the previous study, whose vitamin D content was confirmed as accurate in the Chen-Holick laboratory, a distinguished laboratory with extensive experience in vitamin D analysis.
5) Adherence will confirmed by serum 25(OH)D measurements at baseline and in each year of the study.
6) The accuracy of serum 25(OH)D measurements will be confirmed by participation in the international quality assessment process for vitamin D assays (DEQAS). [19]
7) This study has a similar age range, i.e., at least 60 years, similar to the age group (> 55 years) in which a statistically significant effect of vitamin D and calcium on risk of all cancers combined has been previously shown.[17]
8) There will be frequent follow-up interviews (every 6 months).
9) All cancer cases will be confirmed by physician examination of the medical record using a uniform case definition for each cancer site. Only cancers which have been microscopically confirmed and adjudicated by a central pathologist will be included in the study.
10) Relative homogeneity of the population is likely to reduce the influence of extraneous factors (statistical noise) on the measurement of the effect of the intervention on cancer incidence rates.
_The combination of a very cooperative Midwest population and an experienced research team renders it likely that the proposed project will have excellent adherence for vitamin D. Adherence in our previous study was 86% for vitamin D [17] This is in contrast to vitamin D adherence of about 50% in the WHI.[18] Furthermore, use of separate vitamin D and calcium tablets will allow dropping calcium but not vitamin D. Some previous studies, including the WHI, did not have this flexibility. It is well-accepted that Vitamin D given at 2000 IU/day does not cause adverse health effects,[16] so there is little incentive for dropping it.
_Relationships have already been established with local physicians that will facilitate the proposed study. For example, women often consult their physicians about the wisdom of study participation. In our previous study, local physicians were very supportive.
__ A low rate of attrition is expected, since 87% of women in the previous study completed it.

Thus, we are confident that we will be able to conduct the proposed trial, and we are convinced that it will yield additional critical information on chemoprevention on calcium and vitamin D.