# Primary Cancer and Cardiovascular Disease Prevention: Key Findings from the VITAMIN D and OMEGA-3 TriaL

## 1. Background

**The importance of VITAMIN D**
Although data from laboratory studies, epidemiologic studies, and small clinical trials strongly suggest that vitamin D has a cancer-protective effect, there are no completed or ongoing large randomized trials of high-dose vitamin D supplements for the primary prevention of cancer in a general population. More observational studies, as well as randomized trials if possible, would be beneficial in testing the hypothesis that vitamin D reduces cancer risk. Confirming that vitamin D lowers the risk of cancer incidence or mortality is critical because current health recommendations do not encourage high vitamin D intakes and discourage sun exposure.

Current dietary recommendations are only intended to prevent dangerously low vitamin D levels. To maintain bone health and normal calcium metabolism, the Institute of Medicine recommends 400 IU/d of vitamin D for adults aged 51-70 and 600 IU/d for adults over 70. If the link between better vitamin D status and lower cancer risk is causal, the current levels of intake are likely insufficient.

Vitamin D for cancer prevention remains a challenge, but more research should be prioritized because the potential benefits are substantial. The data on the relationship between vitamin D and CVD is less extensive, but it does support a possible protective effect. It is thus concerning that more than half of the middle-aged and older women in the United States and more than one-third of similarly aged men are vitamin D deficient. Vitamin D for cancer prevention remains a challenge, but more research should be prioritized because the potential benefits are substantial. The data on the relationship between vitamin D and CVD is less extensive, but it does support a possible protective effect. It is thus concerning that more than half of the middle-aged and older women in the United States and more than one-third of similarly aged men are vitamin D deficient. African-American (black) people are especially vulnerable because their skin is less able to synthesize vitamin D in response to solar radiation and they have lower dietary and supplemental vitamin D intakes than whites. Obese people are also at higher risk, owing to decreased bioavailability of this fat-soluble vitamin. Given the aging population and the rising prevalence of obesity. Low vitamin D status is becoming an increasingly serious public health concern.

**The importance of OMEGA-3**
In laboratory and observational studies, marine omega-3 fatty acids have shown significant promise for the primary prevention of CVD; large randomized trials in secondary prevention or high-risk settings have also found benefits. There are, however, no completed or ongoing trials of such supplements for the primary prevention of CVD in a general population selected solely on the basis of age and not on vascular risk factors such as diabetes. According to the American Heart Association's Nutrition Committee, "more studies are needed to confirm and further define the health benefits of omega-3 fatty acid supplements for primary prevention." The conclusion from a 2004 workshop of the ODS and the National Heart, Lung and Blood Institute (NHLBI) Working Group on Future Clinical Research Directions on Omega-3 Fatty Acids and CVD is that "The body of evidence is consistent with the hypothesis that intake of omega-3 fatty acids reduces CVD but a definitive trial is needed".

The VITAL Trial will look at the role of vitamin D and long-chain marine omega-3 polyunsaturated fatty acid supplements (eicosapentaenoic acid [EPA] plus docosahexaenoic acid [DHA]) in cancer and cardiovascular disease prevention (CVD).

## 2. Motivation

According to the CDC statistics, one person dies every 34 seconds in the United States from cardiovascular disease, while cancer is the second leading cause of death nationally. These are ranked as two of the most important health problems facing the nation. Hence some preventions are an important part of the effort to control cancer and cardiovascular disease as both incidences of disease and mortality could be reduced. Vitamin D and Omega-3, which are also referred to as fish oil, are common nutritional supplements that can be accessed in many ways. Some studies suggest that Vitamin D and Omega 3 may prevent or slow cancer or cardiovascular disease development, while some studies oppose that they do not lower the risk for most people. Thus, based on this dataset, we would like to analyze if Vitamin D and Omega-3 do have some impacts on the person-year of different types of cancers and cardiovascular diseases, for which person-year is defined as the estimate of the actual time at risk.

Also, in addition, BMI, which is an estimate of body fat, is a good gauge for the risk of disease that might occur with high body fat in overweight and obese populations. According to research from the American Cancer Society, excess body weight is thought to be responsible for about 11% of cancers in women and about 5% of cancers in men in the United States, as well as about 7% of all cancer deaths. Other diseases such as heart disease, high blood pressure, type 2 diabetes, gallstones, breathing problems, and certain cancers would also be directly affected by excess body fat. At the same time, according to the American Cancer Society, tobacco use remains the leading preventable cause of death in the United States, which is approximately 1 in 5 deaths each year. On average, people who smoke die roughly ten years earlier than those who have never smoked. Cigarette smoking causes 87 percent of lung cancer deaths, which is the highest one. Other than lung cancer, it also causes cancer, heart disease, stroke, lung diseases, diabetes, and chronic obstructive pulmonary disease (COPD), which includes emphysema and chronic bronchitis. Some risk factors for other diseases would also be increased by smoking, such as tuberculosis, certain eye diseases, and problems of the immune system, including rheumatoid arthritis. As BMI and smoking factors are also included in the VITAL dataset, adding to the research question, we would also like to see what roles the BMI and smoking factor would play in the experiment of Vitamin D and Omega-3 with the person-year of different types of cancers and cardiovascular diseases.

## 3. Scientific research question

How would the supplements of Vitamin D or Omega-3 affect the person-year of various types of cancers and cardiovascular, for which person-year is defined as the estimate of the actual time at risk? Additionally, how factors such as BMI and smoking would act in the given experimental trials.

## 4. Data cleaning and exploration

The dataset we use comes from the website of Project Data Sphere and the ClinicalTrials.gov Identifier of the clinical trial case is NCT01169259.

There were 25,871 subjects and 81 characteristics in the data set. 25,871 participants are divided into four groups: 1) daily vitamin D and omega-3; (2) daily vitamin D and omega-3 placebo; (3) daily vitamin D placebo and omega-3; or (4) daily vitamin D placebo and omega-3 placebo. To elaborate a little, the variable “vitdactive” is 1 when the subject was fed by vitamin D and “vitactive” is 0 when the subject took vitamin D placebo. Similarly, the variable “fishoilactive” indicates whether the subject took omega-3 or omega-3 placebo.

The vital dataset contains outcome variables that demonstrate whether each subject developed a specific type of cancer or cardiovascular disease during the years when the trial was conducted, as well as the number of years each subject took to develop this disease. For example, the variable “brca” means whether a subject had breast cancer and the variable “brcayrs” refers to the number of years it took for a subject to develop breast cancer. Similarly, “totmi” indicates whether a subject had myocardial infarction during the years when the trial was conducted and the variable “miyears” shows the years it took for him/her to develop myocardial infarction. We studied a total of eight diseases, including four cancers and four cardiovascular diseases. All variables and labels we used in the trial have been organized in the table below:

**VITAL Study Data Dictionary**

| **Variable** | **SAS Variable Name** | **SAS Variable Label** |
|---|---|---|
| Subject ID | Subject_ID | Assigned Study ID |
| Randomization to active Vitamin D | vitdactive | Vitamin D 1-Active,0-Placebo |
| Randomization to active n−3 fatty acids | fishoilactive | Omega-3 Fatty Acids 1-Active,0-Placebo |
| **Outcomes** | | |
| *Cancer* | | |
| Invasive cancer of any type | malca | Malignant cancer 1-yes,0-no |
| Cancer person-years | malcayrs | Malignant cancer person-years |
| Breast cancer | brca | Malignant breast cancer 1-yes,0-no |
| Breast cancer person-years | brcayrs | Breast cancer person-years |
| Prostate cancer | prca | Malignant prostate cancer 1-yes,0-no |
| Prostate cancer person-years | prcayrs | Prostate cancer person-years |
| Colorectal cancer | colca | Malignant colorectal cancer 1-yes,0-no |

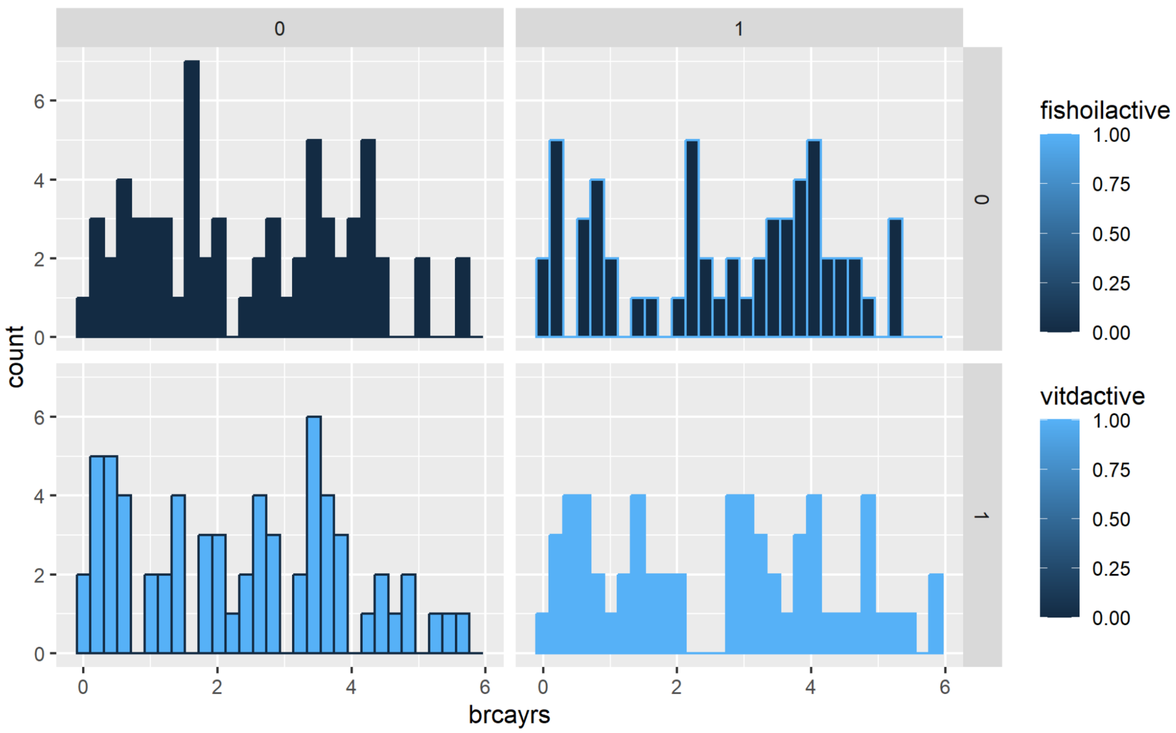
| | | |
|---|---|---|
| Colorectal cancer person-years | colcayrs | Colon cancer person-years |
| | | |
| *Cardiovascular disease (CVD)* | | |
| Major cardiovascular events | majorcvd | Major cardiovascular events 1-yes,0-no |
| Major CVD person-years | majryears | Major cardiovascular events person-years |
| Expanded CVD events | imporcvd | Expanded cardiovascular events 1-yes,0-no |
| Expanded CVD person-years | impryears | Expanded cardiovascular events person-years |
| Myocardial infarction (MI) | totmi | Total myocardial infarction 1-yes,0-no |
| MI person-years | miyears | Myocardial infarction events person-years |
| Stroke | totst | Total stroke 1-yes,0-no |
| Stroke person-years | styears | Stroke person-years |

To begin, we select the columns of all variables shown in the above table. To see if the supplement of vitamin D and omega-3 and the number of years it takes for people to develop the disease are related, we select subjects who had the disease and exclude those who didn't. The subjects with the specific disease are then divided into four groups based on whether they took vitamin D and omega-3 or not, which are labeled (0,0), (0,1), (1,0), and (1,1). Following that, we used the number of years these four-group subjects took to develop a particular disease, counted them separately, and drew a histogram.

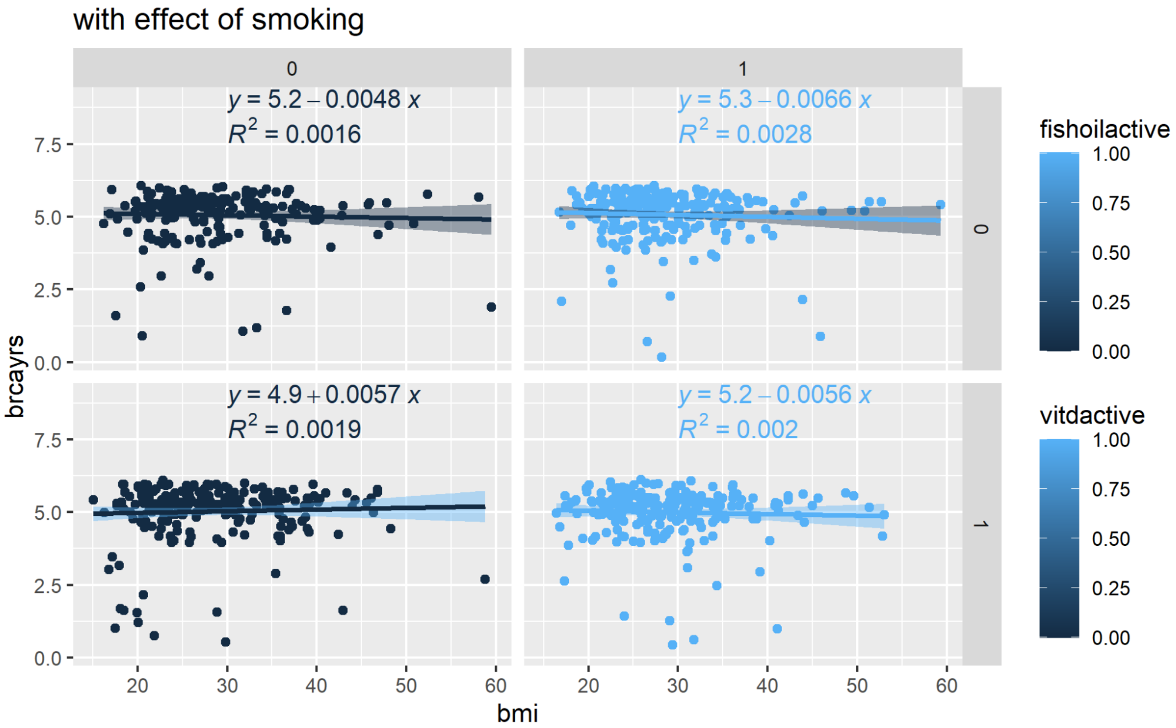
Further, we further added the factors of BMI and smoking to the previous setup. For each group of subjects, we build a linear model, using BMI as the independent variable and the number of years a subject took to develop a specific disease as the dependent variable.

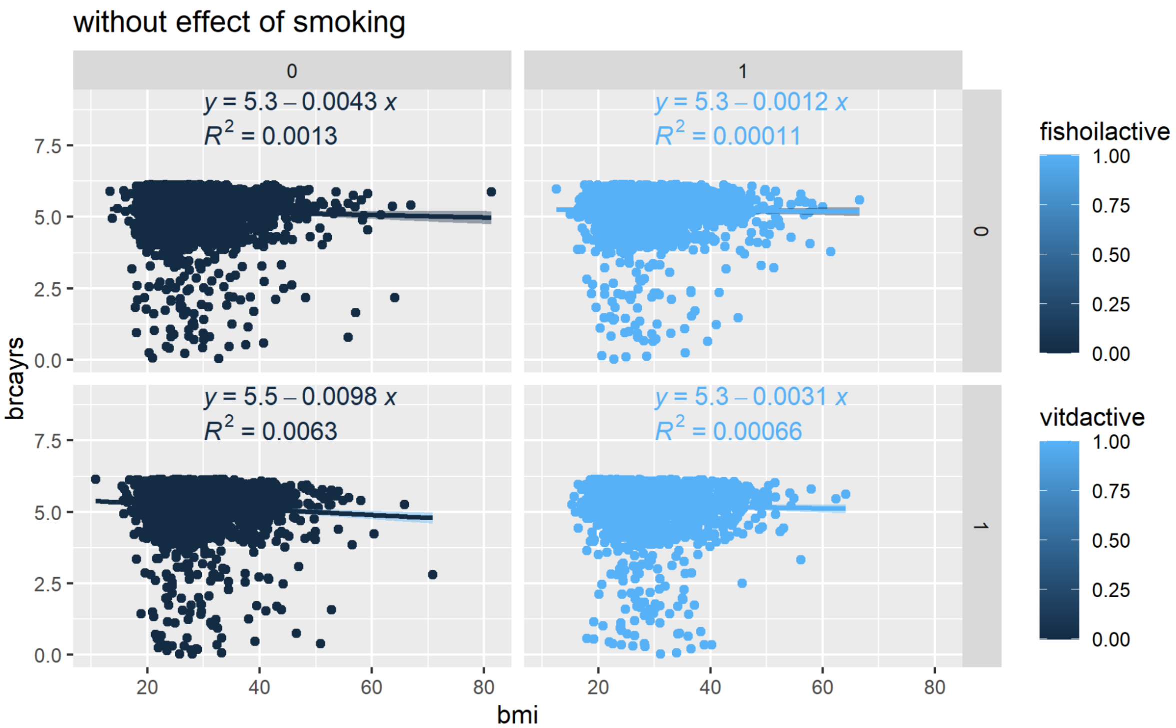
## 5. Analysis and interpretation

The plot below is generated from the vital_hist function with two parameters, one is the dataset vital, and the other one is breast cancer(brca). We studied 8 diseases in total, and the plot here is just taking breast cancer as an example. For other types of diseases, the second parameter in the function could be replaced with the variable names of other diseases, and hence the corresponding plot of other types of diseases would be displayed. For the below graph, there are four subplots with coordinates 0 and 1. 0 represents a placebo, and 1 represents the active intake. For example, (0,0) means taking both vitamin and fish oil as a placebo, and (1,1) means both of them are active. The x-axis is the person-year of breast cancer(the variable name is brcayrs as shown), and the y-axis is the count of the number of cases corresponding to different years at risk of breast cancer. We expected to see some more right-skewed histograms at (1,1) than (0,0), meaning that the effects of intaking vitamin D and fish oil are clearer in reducing the person-year, hence delaying the time at risk and preventing cancer. However, the resulting plot does not show a very clear right-skewed pattern, but it is clear that the plot at (1,1) has a relatively lower average height compared to that at (0,0). Based on the discipline that four groups are divided evenly and randomly, the average number of cases of breast cancer appearing in the group where both fish oil and vitamins are active is lower than the group where both of them are placebo and nonactive. This can also help to infer the positive effects of fish oil and vitamins on reducing the incidence or risk of breast cancer and hence the prevention of breast cancer. The right-skewed pattern is not clear, maybe due to the reason that the intake amount of vitamins and fish oil is not large enough to be clearly displayed in plots. However, too large an intake may also cause health problems for the subjects, so more precise studies are still needed to confirm and further define the health benefits of these supplements as the primary prevention of diseases.




To see how BMI and smoking factors would impact the experiment of Vitamin D and Omega-3, another two functions are written, vital_model_smoke and vital_model_nosmoke. The two functions are similar in their designs, both with two parameters, one is the dataset vital, and the other is the variable name representing one disease. For the plots, we also use breast cancer(brca) as an example. For the other 7 types of diseases, replace the name parameter and the corresponding plots will be displayed. Similar to the previous one, we also divided the data points into four groups, which are (0,0) - both placebo, (1,1) - both active, (1,0) - vitamin active where fish oil placebo, and (0,1) - vitamin placebo where fish oil active, respectively. Instead of histograms, these two functions generate scatter plots, where each point is regarding the patient's person-years of disease versus his/her BMI. In addition to the scatter plots, a model of linear regression is performed on these data points, and the linear relationship for each one is shown together with the linear line. The only difference between the function vital_model_smoke and vital_model_nosmoke is whether the patient is smoking or not. For vital_model_smoke, we filtered all the patients who are smoking, and for vital_model_nosmoke, the filtering criteria is the patients not smoking. From the plot, the intake of vitamins and fish oil does not have a big impact on BMI. This is reasonable because the number of supplements is very small so it won't have a large effect on BMI. Other factors such as eating habits, lifestyle, or genes would have more influence on BMI rather than supplements. However, smoking or not does have some positive impacts on the person-years, which is the time at risk of disease. The R squared in the second plot is generally lower than that in the first plot, meaning that for people who do not smoke, their associations between BMI and time at risk are smaller than those who smoke, so smoking is indeed an associated factor in the prevention of diseases with supplements.



## 6. Conclusion

We can conclude from the above analysis that the right-skewed pattern is unclear, possibly because the amount of vitamins and fish oil consumed is insufficient to be clearly displayed in the plots. However, excessive intake may cause health problems for the subjects, so more precise studies are needed to confirm and further define the health benefits of these supplements as primary disease prevention.

Furthermore, according to the plot, vitamin and fish oil consumption has little effect on BMI. This is reasonable because the amount of supplements is so small that it will not have a significant effect on BMI. Other factors, such as eating habits, lifestyle, or genes, would have a greater impact on BMI than supplements. However, smoking or not smoking has some positive effects on person-years, which is the amount of time spent at risk of disease. The R squared in the second plot is generally lower than in the first, indicating that for people who do not smoke, the associations between BMI and time at risk are smaller than for those who do, indicating that smoking is indeed an associated factor in the prevention of diseases with supplements.