**May 2018**

**Case-cohort study**

**Selection of a case-cohort design:**
A baseline blood sample was collected on almost all of the SELECT participants with a pre-defined study objective of analyzing baseline biomarkers. Most of the hypotheses based on baseline biomarkers can be thought of as taking two forms.

1. Is a baseline measure (e.g., presence or absence of a SNP, serum selenium concentration) associated with risk of prostate cancer?

2. Is the association of study treatment with prostate cancer modified by the participant's baseline measure (that is, is there a treatment by factor interaction)?

It is not necessary to analyze serum from all participants in the SELECT trial in order to make valid inferences and have good statistical power to evaluate these trial objectives. Some type of sampling is needed whereby we analyze incident cases and a subset of non-cases. A case-cohort design is proposed because in addition to prostate cancer, SELECT is interested in evaluating other endpoints such as Alzheimer's disease, respiratory function, macular degeneration, and colorectal cancer. By choosing a cohort design, the same cohort can be used to analyze associations with the other endpoints of interest. Other additional work needed would be the biologic analysis of the respective "cases" that were not selected in the cohort.

Benefits of the case-cohort include the fact that the cohort can be used as a contrast group for multiple outcomes (ancillary studies), the cohort can be identified at the beginning of the trial so easy to plan and implement and finally, laboratory analysis of biological specimens can be completed and data ready for analysis at study completion.

Equal proportion of men from each of four age categories for African-American and non-African American men (< 55 (African—American only), 55-59, 60-64, 65-60, ≥ 70) were selected for the analysis cohort. These factors were chosen as they were felt to be of importance not just for prostate cancer but for other potential disease of interest including the endpoints of the various ancillary studies. African-American men were oversampled to increase power in this important subgroup. Sampling ratios were 1:3 for African Americans and 1:1.5 for all others.

For implementation of the design, for each age group x race stratum (9 cells), we generated a random list of all subjects from SELECT that fell into that category. Then, starting in 2005, at each year, we determined the number of incident prostate cancers in each stratum and choose the corresponding 1.5 x (or 3.0 x for AA) corresponding cohort samples. Samples from the new incident cases and selected sub-cohort were processed annually. Laboratories were blinded to treatment and case status.

The final selection of cases was in June, 2009. At that time there were 1856 cases and 3188 controls for a total sample size of 5044. Bloods and toenails were processed and DNA extracted on an annual basis (incident cases and an appropriate number from the cohort) to minimize any potential bias of sample degradation over time. As of June 1, 2010, the DNA extraction for all men with an available sample is complete (n=4882), the analysis of the available toenail samples for selenium is complete (n=4748) and the plasma analyses are approximately 60% complete.

Of the 5044 men who are in the case-cohort, 3081 (1070 cases) have had the mid-study blood draw so they have a post-baseline sample. Additionally, 377 (126 cases) of the 5044 men are in the adherence cohort and have more than one post-baseline sample. None of these plasma samples have been analyzed.

**Power:**

Table 1: Minimally Detectable Hazard Ratio (HR) for Prostate Cancer, Varying Marker Prevalence Among Case-Cohorts, Assuming 1-β=85%, α=0.05

| | | **Prevalence of Marker or SNP** | | | | |
|---|---|---|---|---|---|---|
| | **Cases / Cohort*** | **2%** | **5%** | **10%** | **30%** | **50%** |
| HR for **case-cohort** | 1856 / 3,404 | 1.74 | 1.43 | 1.29 | 1.19 | 1.17 |
| HR for **full clinical trial** | 1856 / 30,762 | 1.58 | 1.34 | 1.23 | 1.15 | 1.13 |
| HR for **AA men in case-cohort** | 263/847 | 3.80 | 2.36 | 1.85 | 1.49 | 1.45 |
| HR for **AA men** in **full trial** | 263/3,771 | 3.30 | 2.15 | 1.73 | 1.44 | 1.39 |

Table 2: Minimally Detectable Hazard Ratios for Other Cancer Endpoints Using DNA Samples from **All** of SELECT

| | | **Prevalence of Marker or SNP** | | | | |
|---|---|---|---|---|---|---|
| | **Cases / Cohort** | **2%** | **5%** | **10%** | **30%** | **50%** |
| Lung cancer | 353 / 30,762 | 2.80 | 1.92 | 1.62 | 1.37 | 1.33 |
| Colorectal cancer | 304 / 30,762 | 3.00 | 2.05 | 1.67 | 1.40 | 1.36 |
| Pancreatic cancer | 106 / 30,762 | 6.50 | 3.35 | 2.38 | 1.77 | 1.69 |