Date: April 8, 2016

To: P and P Committee/ Reviewers

From: Bette Caan, Ross Prentice, Aaron Aragaki

Re: Guidelines for cancer mortality and survival analyses generated from WHI data.

In an attempt to standardize WHI papers using cancer outcomes that include death we have provided the following guidelines:

LILAC Papers

With the start of data available from LILAC you will likely see a number of papers examining effects of exposures on cancer survival.

- These papers are intended to examine the effect of an exposure on survival once a person is diagnosed with cancer.
- The time scale for follow-up should start at cancer diagnosis. The exposure/covariate histories considered should be relevant to the time of diagnosis and post-diagnosis period. This may be a particularly important consideration if the exposures/covariates are modifiable risk factors that can change over a woman's follow-up period.
- In most situations, the Cox proportional hazards model for the survival analysis should include stratification on age at cancer diagnosis in 5-year intervals (i.e., use a separate baseline hazard function for each 5-year group). In the presence of sparse data, larger groups (e.g., 10-year intervals) may be used.
- Note that if the exposure is constant over the woman's follow-up period (e.g. reproductive history) it may even be useful to examine exposures that occurred many years previous to diagnosis relative to post-diagnosis events (with the follow-up time clock starting at diagnosis). However, the interpretation of post-diagnosis associations may be affected if the exposure is related to cancer incidence, and it is important to ensure that there is adequate adjustment for confounding. The potential for selection bias (in the case where exposure is related to incidence) should be acknowledged in the discussion section.
- Treatment data on a portion of cases is available for use by the CCC analysts who can re-run analyses performed by outside investigators including treatment; these papers should include treatment data, wherever possible, even if it is only in a sensitivity analyses to demonstrate that treatment does or does not alter the observed associations. Currently treatment data may be used by CCC analysts to perform sensitivity analyses for:
    - Breast cancer cases in Medicare + non-Medicare breast cancer cases diagnosed 2007-12
    - Colorectal cancer cases in Medicare

Cancer Incidence and Death papers

To date, for a variety of exposures measured at baseline, we have been publishing a number of papers on cancer incidence and what we have to date been using either cancer-specific death or cancer-

specific mortality in our papers. The time origin for follow-up is the date of WHI enrollment or randomization. While we will continue this type of analysis, we need to make sure that our readers understand the difference between this type of death analyses and survival analyses referred to above since results can differ markedly. One good way to separate out the two different types of analyses is to use different terminology for the two types of cancer outcome.

- We are proposing that the when the time clock starts at enrollment/randomization authors use cancer-specific death(such as breast-cancer death) and when the time clock starts at cancer diagnosis authors use cancer-specific survival or mortality (such as breast cancer survival or breast cancer mortality).
- The same pertains to the terminology when using overall death (time from enrollment) and overall survival/ mortality (time from diagnosis).
- When overall death is used and includes any cause of death, it should not be restricted to only those who have cancer since everyone in the cohort, regardless of a cancer diagnosis is at risk of dying.
- In these type of analyses treatment has not been traditionally included which is OK because it involves the entire cohort and treatment is not available to those who are not diagnosed with cancer.
- Some ways to alert the reader that the time clock starts from enrollment or randomization is to 1) include the time scale in the abstract 2) put a sentence in the abstract recommending that future work examine effects of post-diagnosis exposure on survival and 3) include two curves in the paper; one demonstrating the effect when you begin the clock at enrollment on death and the other when you begin the clock at diagnosis on survival. For these displays, a cumulative hazard function (curves begin at zero on the y-axis) could be used to summarize death, and a survivorship function (curves begin at unity on the y-axis) to summarize survival.