6.0 STATISTICAL ANALYSIS

6.1 Primary and secondary endpoints

Endpoints were selected with the input of patient stakeholders in the general categories of: 1) breast cancer and health-related clinical outcomes including pulmonary, cardiovascular, other cancer, and
hematologic disease), and 2) Patient Reported Outcomes (PRO) which will include comprehensive domains of QOL and psychosocial outcomes, with a primary focus on pain and physical symptoms.

6.2 Sample size and statistical power

Importantly, published data indicate that patients undergoing lumpectomy with radiation report comparable levels of pain and sensory disturbances as those undergoing mastectomy.[56] Thus, these treatment groups will be grouped together in the initial analysis, then a stratified analysis will be performed. Preliminary data among 582 women undergoing either lumpectomy or mastectomy indicate that the mean persistent pain intensity after breast surgery is 2.61 (DS 4.24).[55] Given a sample size of 900 patients in the GCC group and 300 in the AS group, we will be able to detect with 90% power at a level of 0.05 (2-sided) a difference in pain intensity score of 2 (from 2.61 to 0.61), which has been deemed to be a clinically meaningful difference.

Based upon the number of unique cases diagnosed with DCIS at each site between 2010 and 2013, we estimate that at all 6 study sites combined, approximately 1500 women will have been diagnosed with DCIS between January 1, 2012 and June 30, 2017 and will be eligible to participate. Our response rate in prior survey studies has been approximately 60% using the approach we will adapt to the current study; thus, we anticipate that about 900 women with DCIS will be accrued and will complete the survey across all study sites. These patients will be matched 3:1 to participants diagnosed with ADH or LCIS/ALH undergoing active surveillance. DCIS participants will be matched 3:1 to ADH/LCIS/ALH participants, based on matching criteria of age within 5 years, year of diagnosis, and study site. Additional matching criteria such as race, income, highest education level attained, and family history of breast cancer will be considered based upon stakeholder input during the protocol preparation process.

6.3 *Analysis Plan: General Approach.*

We will conduct descriptive analyses to profile the samples, including examination of proportions, means, and medians, as well as estimates of variability such as standard errors, ranges, and confidence intervals. We will identify outlying observations. Continuous data distributions will be evaluated for appropriateness of scale, and normal score transformations will be used where

appropriate. Appropriate summary statistics, histograms, scatter plots, or one- or two-way contingency tables for women of each ethnic group will be computed. These analyses will be followed by bivariate analyses to examine the pairwise relationships among variables under examination. Demographic data will be compared for the respondents and non-respondents using chi-square or t-tests to assess selection bias. In addition to baseline disease and socio-demographic data, information relating to other treatment (e.g., endocrine treatment, radiation) as well as co-morbidities will be collected, allowing us to control for potential confounding by these factors. We will also analyze the decision-making, DCIS knowledge, and risk perception data at baseline and compare these domains between the GCC and ADH/LCIS/ALH/AS groups. We will also systematically collect and evaluate and report on the upstaging rates and associated details from initial diagnosis of DCIS.

6.4 ***Management of missing data:*** As with all studies there is a risk of biased statistical inference with missing data and we will adhere closely to standards put forth by the National Research Council. [77] We will include details about missing patient data in all tables in subsequent analyses, and include a comparison of baseline characteristics of patients with and without missing data. We will carefully consider whether data missingness is random or whether missingness may be attributable to a certain factor and therefore not random (e.g., older women may be less likely to answer questions). We will assess non-response bias by comparing available characteristics in both patients and providers (e.g., practice type) among responders and non-responders.