Title: Statistical Analysis of Breast Cancer Survival in the US

Background: Breast cancer deaths are decreasing in the United States overall, but disparities still exist. Despite progress, it is found that non-Hispanic Black women had a 41% higher death rate from breast cancer than non-Hispanic White women. The use of statistical methods is an imperative tool in breast cancer survival data analysis. The purpose of this study was to identify the best statistical probability model using the Bayesian method to predict future survival times for the black non-Hispanic female breast cancer patients in the U.S.

Materials and Methods: We used a stratified random sample of black non-Hispanic female breast cancer patient data from the Surveillance Epidemiology and End Results (SEER) database. Survival analysis was performed using Kaplan-Meier and Cox proportional regression methods. Four advanced types of statistical models, Exponentiated Exponential (EE), Beta Generalized Exponential (BGE), Exponentiated Weibull (EW), and Beta Inverse Weibull (BIW) were utilized for data analysis.

Results: We identified the highest number of black non-Hispanic female breast cancer patients in Michigan and the lowest in Hawaii. The mean (SD), of age at diagnosis (years) was 58.3 (14.43). The mean (SD), of survival time (months) for black non- Hispanic females was 66.8 (30.20). Non-Hispanic blacks had a significantly increased risk of death compared to Black Hispanics (Hazard ratio: 1.96, 95%CI: 1.51–2.54). Compared to other statistical probability models, we found that the exponentiated Weibull model better fits for the survival times.