**Supplementary Figure 1: Distribution of HER2-positive and HER2-negative tumors by age at diagnosis before and after the imputation, SEER-18 excluding Alaska.**

Supplementary Figure 1 shows original (solid blue line) and imputed (dashed blue line) distributions for HER2 status by age at diagnosis based on one of the 25 imputed datasets (similar results were found when we repeated the analysis with a second imputed dataset). The frequency distribution of HER2+ and HER2- tumors are shown before and after imputation by age at diagnosis in panel 1A

and panel 1B, respectively. Percent of HER2+ and HER2- tumors before and after imputation by age at diagnosis are shown in panel 1C and panel 1D, respectively.

**Supplementary Figure 2: Four-year breast cancer-specific survival by stage and molecular subtypes before and after imputation, SEER-18 excluding Alaska.**

Breast cancer-specific survival curves using original (solid lines) and imputed (dashed lines) data by stage at diagnosis and molecular subtypes and are shown in panel 2A) Among Stage I disease, 2B) Among Stage II disease, 2C) Among Stage III disease, and 2D) Among Stage IV disease: HR+/HER2- (blue solid and dashed lines), HR+/HER2+ (red solid and dashed lines), HR-/HER2+ (green solid and dashed lines), triple negative (black solid and dashed lines), and unknown (grey solid line). Note that after imputation, there are no unknown subtypes reaming in the dataset.