Long-Term Survival Prediction in Early Breast Cancer : A Machine Learning Approach with Random Survival Forest

Hyuna Yoon[1,2,3], Sola Han[1], Hae Sun Suh[2,3,4], Chanhyun Park[1]

1. College of Pharmacy, The University of Texas at Austin; 2. College of Pharmacy, Kyung Hee University; 3. Institute of Regulatory Innovation through Science, Kyung Hee University, Seoul, Republic of Korea; 4. Department of Regulatory Science, Graduate School, Kyung Hee University.

MSR49

BACKGROUND

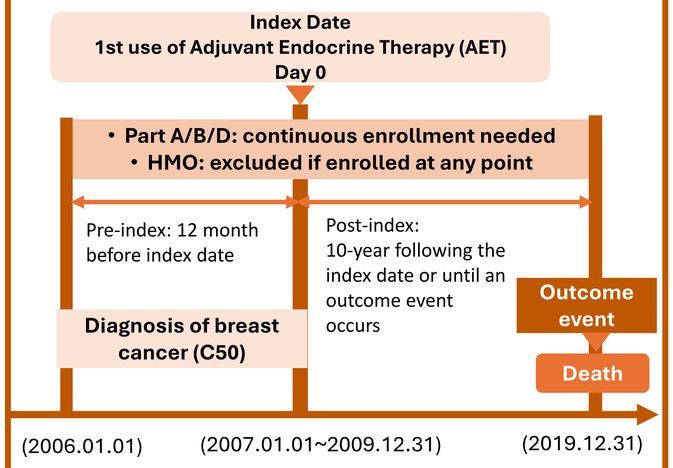
- Breast cancer (BC) is the most common cancer among women, and it has the second-highest mortality rate among all cancers in women in the US.[1]
- Regarding the causes of death among patients with breast cancer, the primary cause is breast cancer itself, and among non-cancer related causes of death, cardiovascular disease is the most common.[2,3]
- Predicting breast cancer survival curves can help improve patient outcomes.
- The Random Survival Forest (RSF) algorithm showed good predictive performance in predicting breast cancer patient's death.[4,5]

OBJECTIVES

- To develop survival prediction models for all-cause, BC-caused, and cardiovascular disease (CVD)-caused mortality in older women with hormone receptor-positive (HR+) early BC.
- To identify key prognostic factors for mortality.

METHODS

- **Study Design**: A retrospective cohort study
- **Data Source:** SEER-Medicare
- **Study Population**
  - Female
  - Early-stage breast cancer
  - Had primary surgery for breast cancer including breast conserving surgery
  - Age 66 years old and older
  - Hormone receptor positive



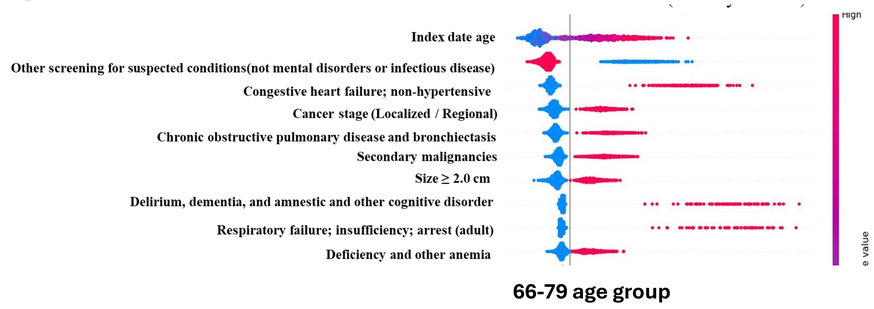
- **Model Development**: Random Survival Forest
- **Training set (75%) / Test set (25%)**
- **Variables:** Demographic (age, race), Cancer-related (tumor size, SEER summary stage), Diagnosis-based comorbidity defined using Clinical Classification Software (CCS) (285 categories)
- **Feature Selection:** LASSO method in Cox Model
- **Hyper-parameter tuning**
  - Number of estimators: (10, 100, 500, 1000)
  - Minimum of samples split: (2, 3, 5, 10, 15, 20)
  - Minimum of samples leaf: (1, 3, 4, 10, 15, 20)

RESULTS

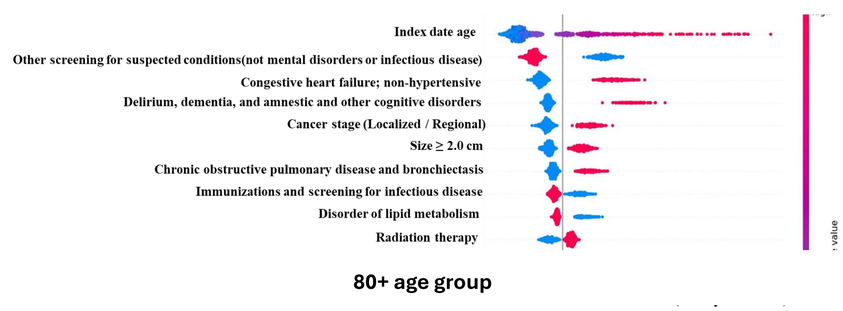
Figure 1. Model Performance

	Old (66-79 years old) Patient			Oldest old (≥ 80 years old) Patient		
	All cause death	BC-related death	CVD related death	All cause death	BC-related death	CVD related death
C-index	0.719	0.789	0.742	0.693	0.756	0.707
Mean AUROC	0.771	0.811	0.737	0.763	0.805	0.765
IBS	0.109	0.047	0.016	0.154	0.073	0.054

Figure 2. SHAP Plots for All-Cause Death

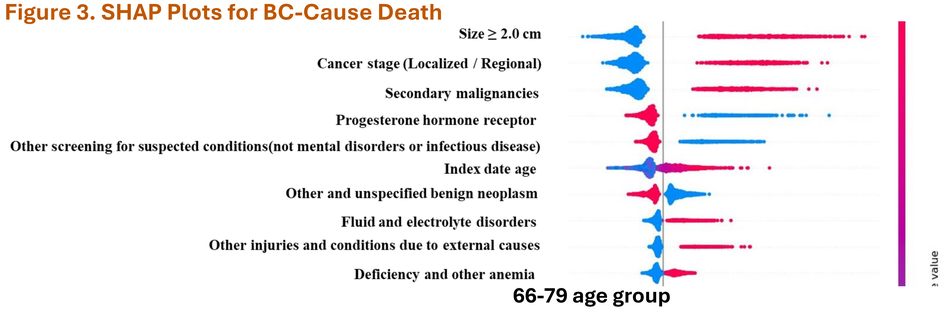


66-79 age group

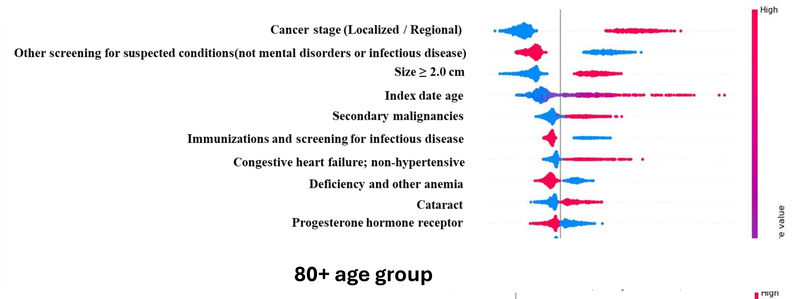


80+ age group

Figure 3. SHAP Plots for BC-Cause Death

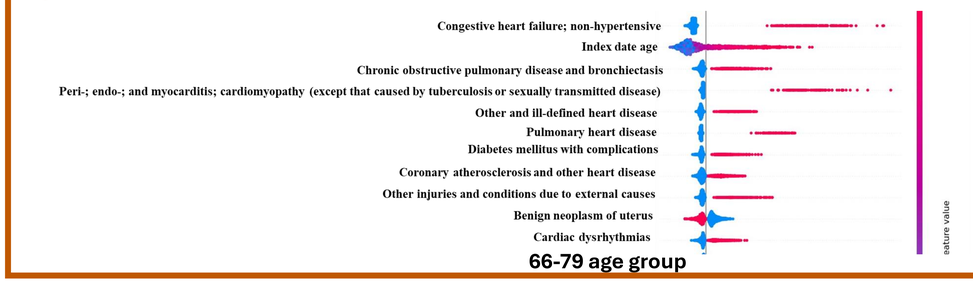


66-79 age group

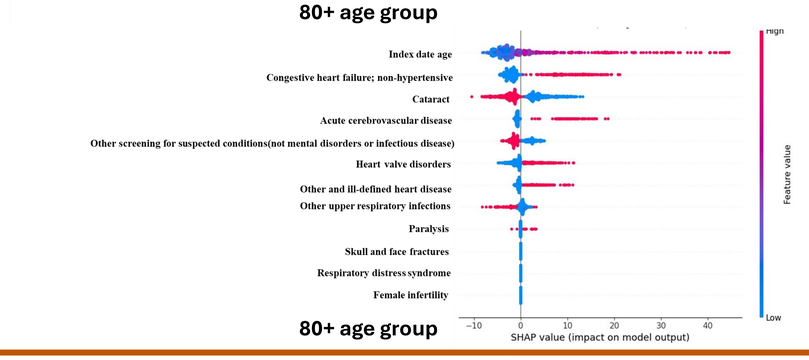


80+ age group

Figure 4. SHAP Plots for CVD-Cause Death



66-79 age group



80+ age group

CONCLUSION

Strengths

- This is the study that developed 10-year long-term survival prediction models of the older patients with breast cancer for the U.S. population.
- This study included patient's comorbidities, which are not included in the prior studies, as the model predictors.
- This study explained the developed model using SHAP and identified the factors that influence patients' death.

Conclusion

- We developed survival prediction models for older women with breast cancer which showed acceptable performances using random forest algorithm.
- Age, screenings, and congestive heart failure were key predictors of mortality of all cause death regardless of age.
- Cancer characteristics such as tumor size, cancer stage, and presence of secondary malignancies were the key factors for BC-caused mortality.
- Congestive heart failure, heart valve disorder, and other ill-defined heart diseases were the key predictors for the CVD-caused death.

REFERENCE

1. Siegel, R. L. et al., *CA Cancer J Clin*. 2023; 73(1), 17-48.
2. Afifi, A.M. et al., Cancer. 2020; 126(7), 1559-1567.
3. Reddy, P. et al., CA Cancer J Clin. 2017;8(4) 308-314 .
4. Najafi-Vosough, R. et al., J Prev Med Hyg. 2022; 63(2).
5. Rahman, M. S. et al., ICECET. 2023.

Acknowledgement This research was supported by a grant (21153MFDS601) from Ministry of Food and Drug Safety in 2024.

Contact: hyuna.yoon@austin.utexas.edu