

# Cardiovascular Risk Stratification Using Unsupervised Machine Learning in Breast Cancer Patients Initiating Anthracycline-Based Treatments

Udim Damachi, MS, PhD<sup>1</sup>, Manu Murali Mysore, MD<sup>2</sup>, Mathangi Gopalakrishnan, MS, PhD<sup>1</sup>, Susan dosReis, PhD<sup>1</sup>, Eberechukwu Onukwugha, MSc, PhD<sup>1</sup>, Wendy Camelo Castillo, MD, MSc, PhD<sup>1</sup>

<sup>1</sup>School of Pharmacy, University of Maryland, Baltimore, MD, USA, <sup>2</sup>School of Medicine, University of Maryland, Baltimore, MD, USA.

## BACKGROUND

- Approximately 39% of women with breast cancer have more than one cardiovascular risk factor (CVRF) and conditions related to cardiovascular disease (CVD) before receipt of anthracycline treatment.
- Older age, concurrent cardiotoxic treatments, and pre-existing CVD increase the risk of anthracycline-induced cardiotoxicity (AIC).
- Concurrent CVRF reduces cardiac resilience, leading to differing risk and greater severity of AIC, especially with multiple coexisting risk factors.

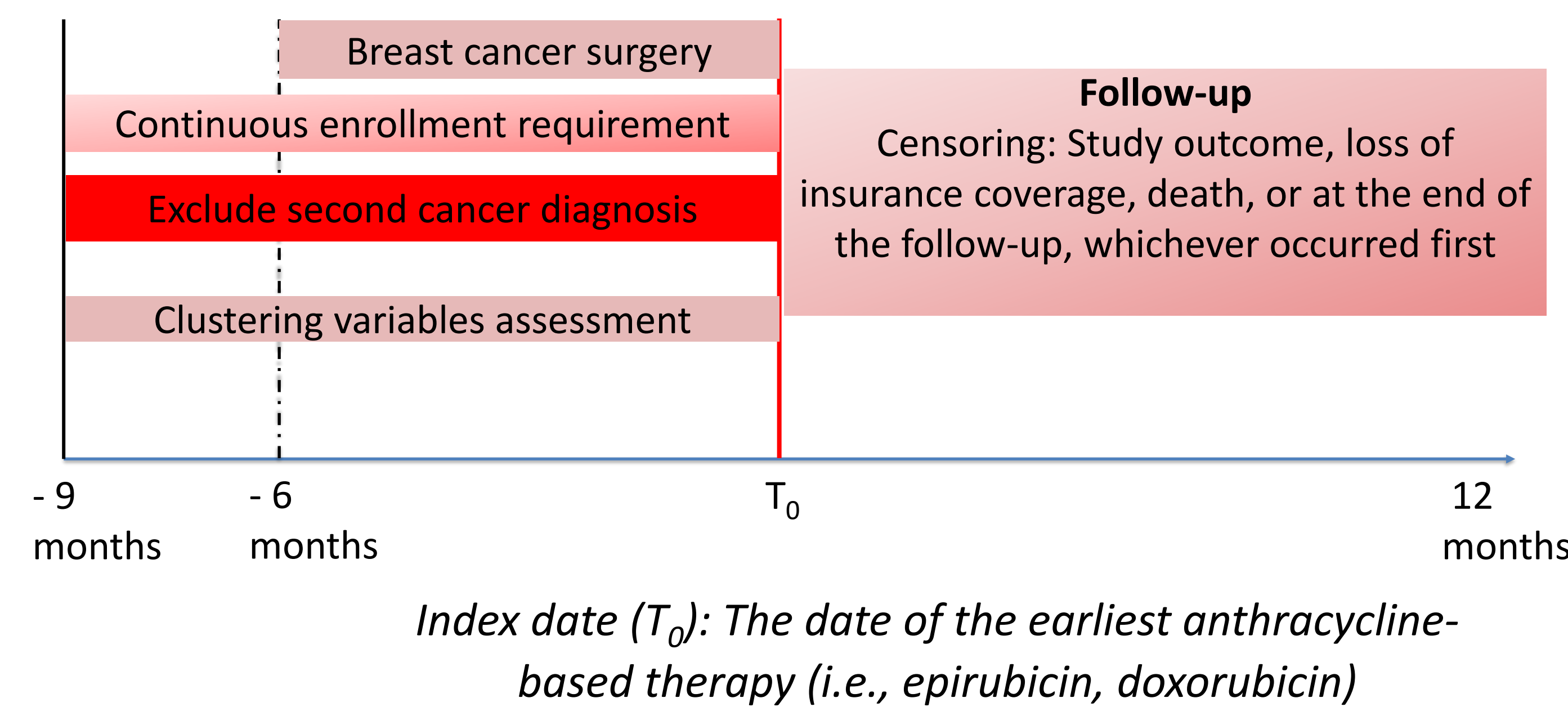
## OBJECTIVES

- To explore whether unsupervised machine learning could identify latent subgroups defined by distinct CVRF profiles.
- To assess the association between CVRF profiles and AIC-related outcomes in women with breast cancer initiating adjuvant anthracycline treatment.

## METHODS

- **Design and Data source:** Retrospective cohort study using 25% random sample of 2006-2024 IQVIA PharMetrics® Plus US health plan claims data.
- **Study Cohort:** Women aged 18-84 years with a claim for a breast cancer diagnosis and breast cancer surgery six months before anthracycline treatment initiation comprised the cohort. The index date ( $T_0$ ) was the start of adjuvant anthracycline treatment (See Fig.1).
- **Clustering Method:** K-prototype Clustering for continuous and categorical data.
- **Optimal Cluster Selection:** Elbow method and Silhouette Width criteria.
- **Variables for Clustering:** Age, cancer treatments (radiotherapy, HER2, cyclophosphamide, taxanes, fluorouracil), CVRF (diabetes, hypertension, obesity, hyperlipidemia), CV disease conditions, and cardio-protective medication use (antihypertensives, antidiabetics, statins).
- **Study Outcomes:** AIC-related hospitalization/emergency department (ED) visit related to heart failure or cardiac arrhythmia.
- **Exposures:** Patient-identified clusters using K-prototype clustering.
- **Statistical Analysis:** ANOVA and  $\chi^2$ , Fisher's exact test with Monte Carlo simulation tests evaluated characteristics across identified clusters.
- We examined the risk of AIC-related hospitalization/ED visit within each cluster using a Cox proportional hazard regression model.

Fig.1 Study design Illustration

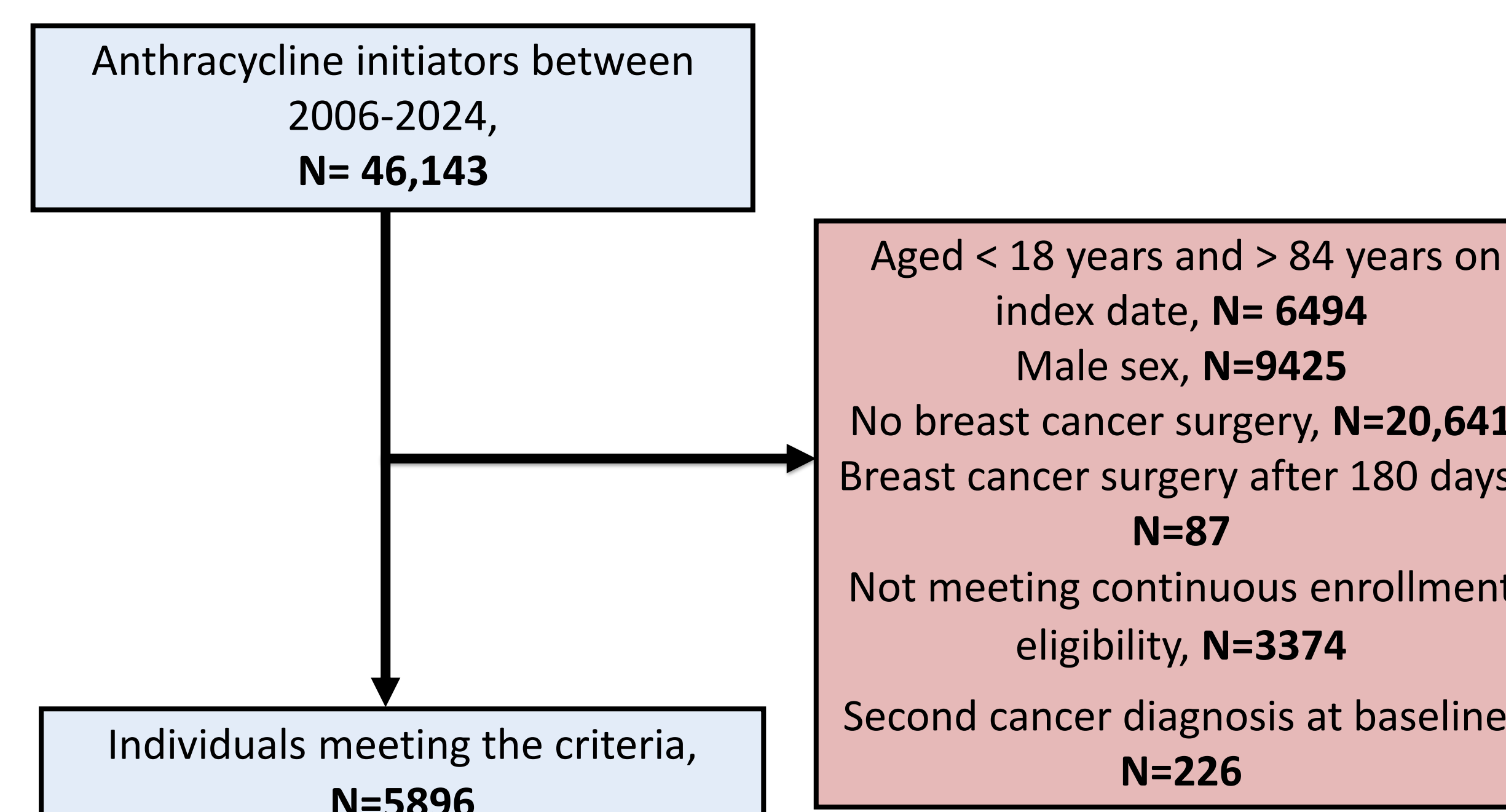


## RESULTS

### Study Cohort

- 5,896 women with a mean age: 52.4 years (SD 9.7) (see Fig. 2)
- 6.1% had established CVD
- 44.4% had  $\geq 1$  CVRF

Fig. 2 Cohort Selection Diagram



### Patient Clusters

- We identified four clusters (See Fig.3)
- **Cluster 1:** Mostly aged 50-59 years with a low CVD prevalence.
- **Cluster 2:** Highest prevalence of radiotherapy (43.6%) and CVRF - hyperlipidemia (93.6%), diabetes (32.6%), and obesity (26.4%).
- **Cluster 3:** Mostly younger (mean: 43.4 years old) with a low CVD prevalence.
- **Cluster 4:** Mostly aged  $\geq 60$  years with hypertension and antihypertensive therapy.

Fig 3. Proportion of Baseline Characteristics by Clusters

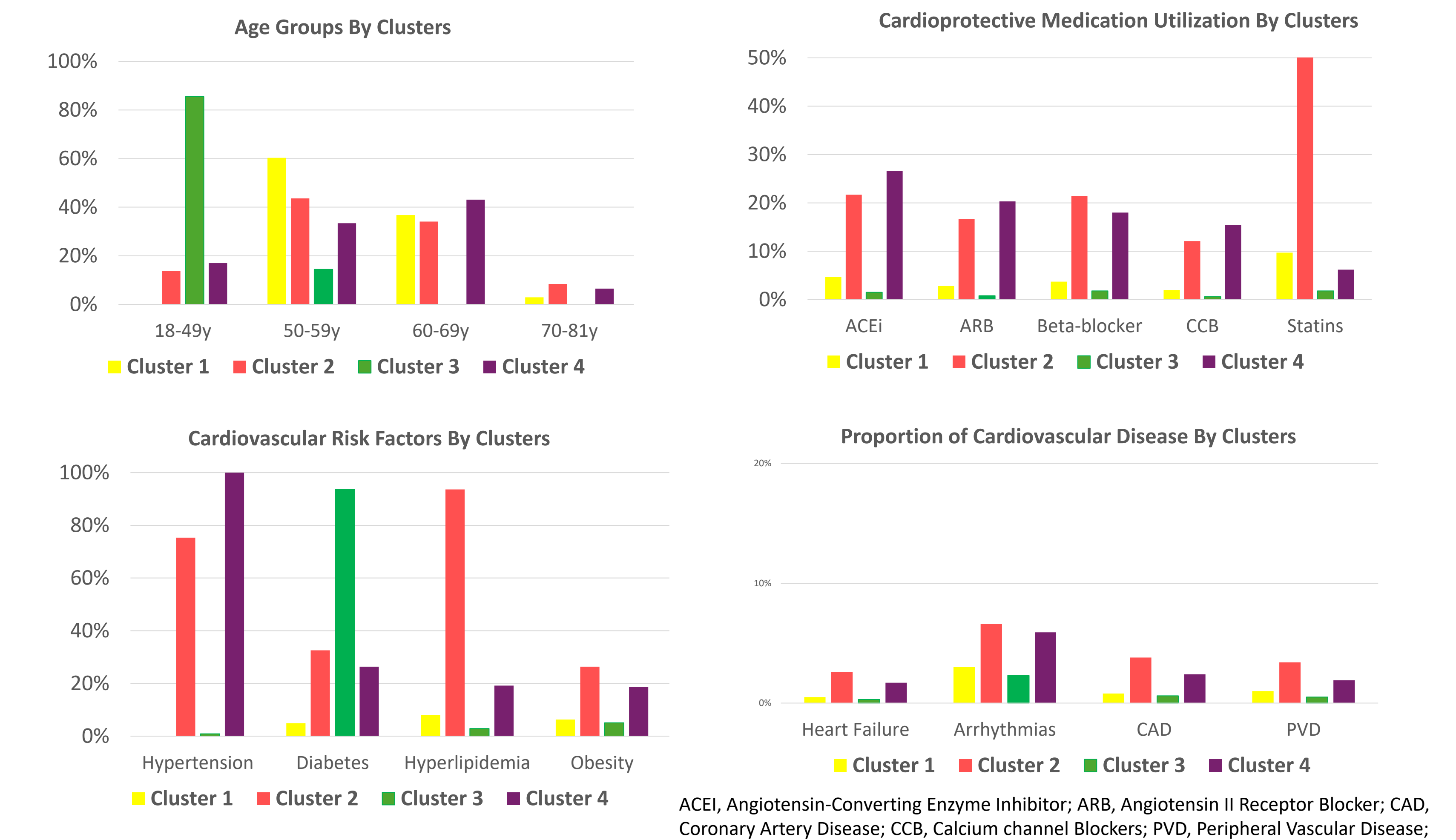


Table 1. Cox Proportional Hazard Models of Clusters of AIC risk factors and Risk of Anthracycline-induced Cardiotoxicity

Clusters	Person-years	No. of Events	Incidence per 10,000 Person Years	Unadjusted HR (95% CI)
Cluster 3	770,818	14	0.18	1
Cluster 1	619,630	25	0.40	2.22 (1.15-4.26) **
Cluster 2	182,830	14	0.77	4.20 (2.00-8.80) **
Cluster 4	340,326	15	0.44	2.42 (1.17-5.02) **

Abbreviations: HR, Hazard ratios; No., Number, \*\* statistically significant

Reference group: Cluster 3 was selected as the reference group because it represented the subgroup with the lowest CV risk profile at baseline.  
Events: Anthracycline-induced cardiotoxicity-related (i.e., heart failure or cardiac arrhythmias) hospitalization or Emergency department visits.

## CONCLUSION

- Using real-world data and K-prototype clustering, we identified distinct CVRF subgroups among women initiating ABT.
- Cluster 2 exhibited the highest risk of AIC-related hospitalization/ED driven by the combined influence of age and pre-existing CVRF and conditions.

## FUNDING

- This work was supported by the 2024 PhRMA Foundation Predoctoral Fellowship in Value Assessment and Health Outcomes.