Population Data Analysis

Analysis of Population Data

Goal: Investigate treatment outcome in stroke populations and identify risk factors

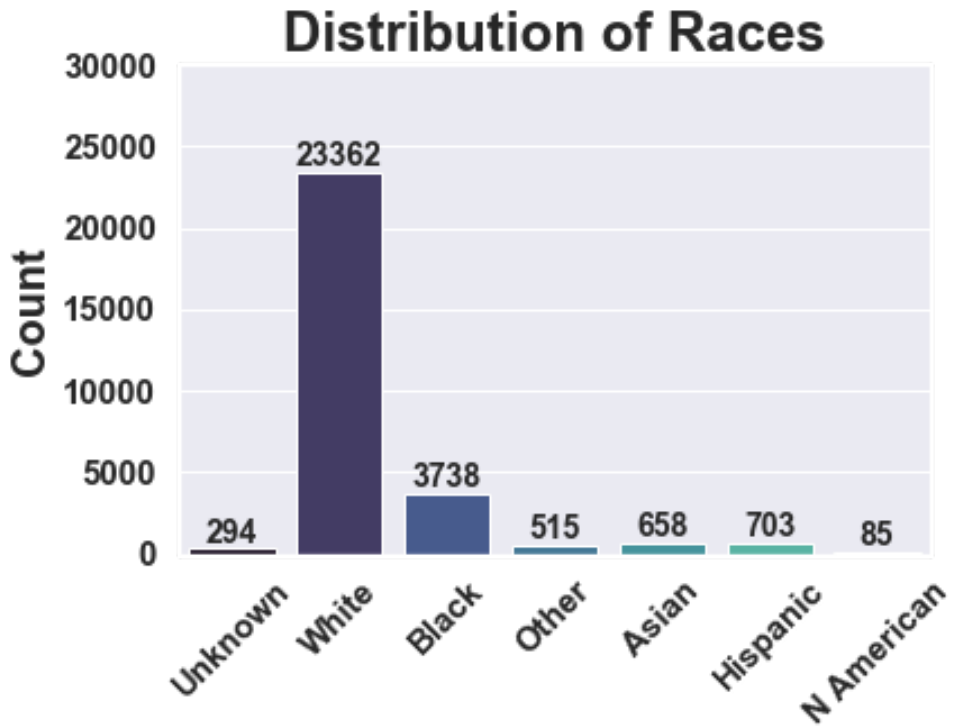
- *The MEDPAR CMS Dataset*
- From the Center for Medicare and Medicaid (CMS)
- Described as the Medicare Provider Analysis and Review (MEDPAR)
- Contains info for 100% of Medicare beneficiaries using hospital inpatient services for acute ischemic stroke (AIS)

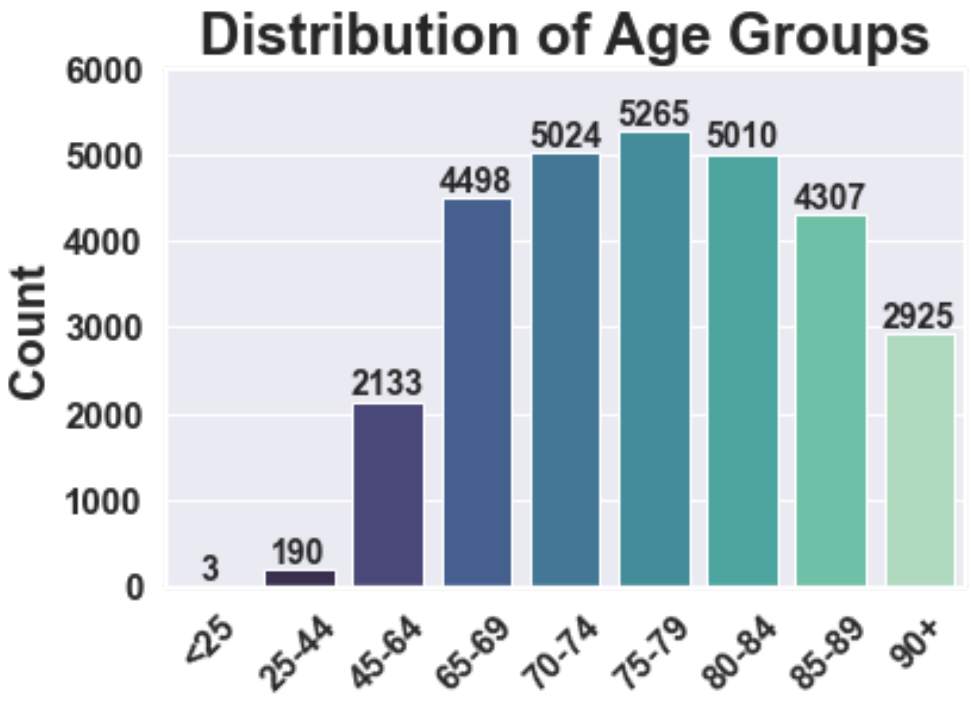
- Prior to data cleaning and pre-processing…
  - ***n* = 29,355 patients**
  - Original data included fields such as:
    - patient demographics
    - summary insurance charges and Medicare reimbursements
    - duration of stay, organized by hospital unit
    - diagnostic codes for admission and duration of stay

Data Cleaning

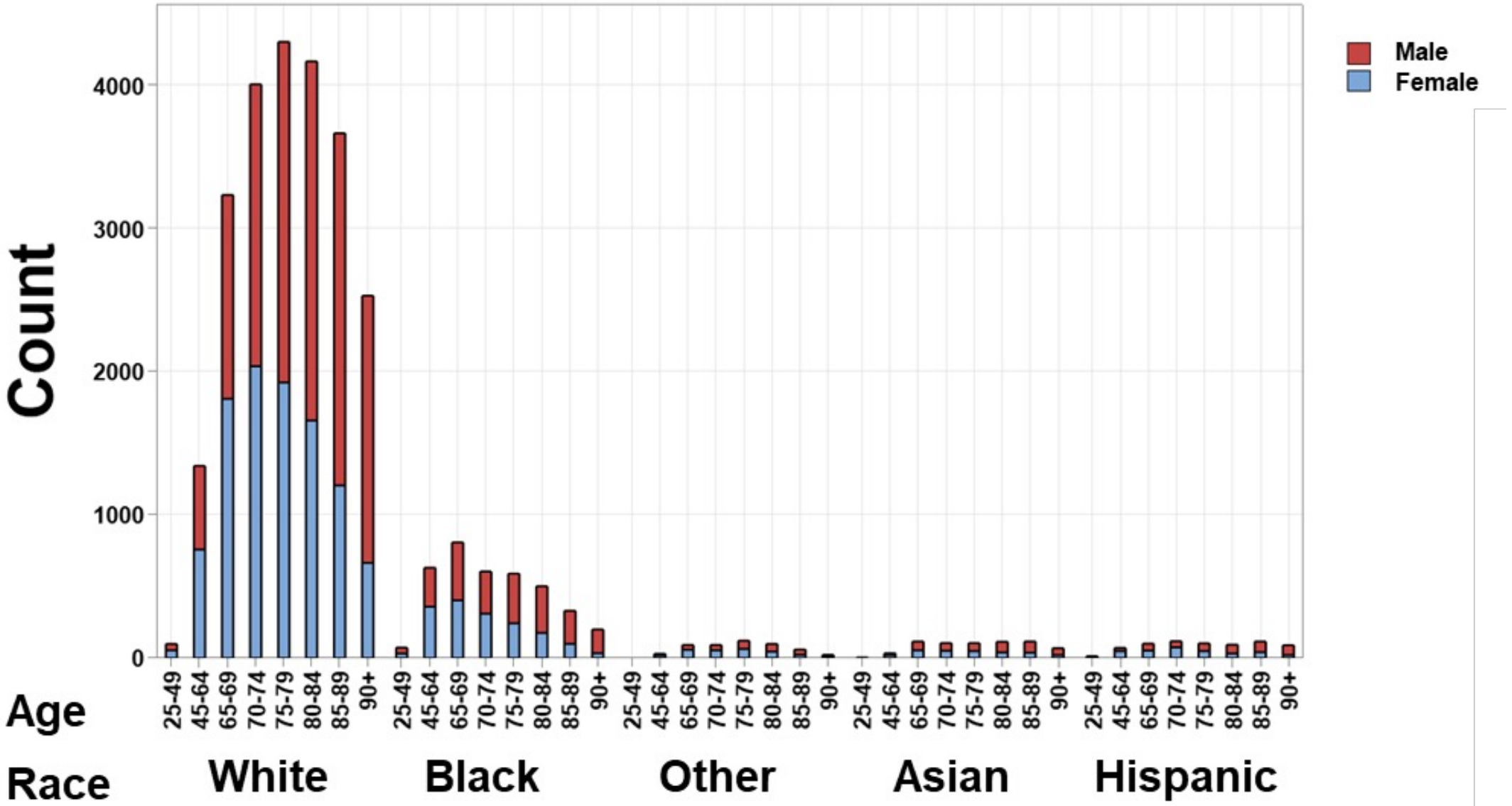
- Filtering based on the following criteria:
  - Demographic subpolns > 100 members
  - Demographic subpopIns w/ unknown status
  - Completeness and relevance of features
- Deduplication
- Outlier Removal

→ **Final *n* = 28970 patients**





Demographic Distribution of Patient Data



Data Transformation and Feature Reduction

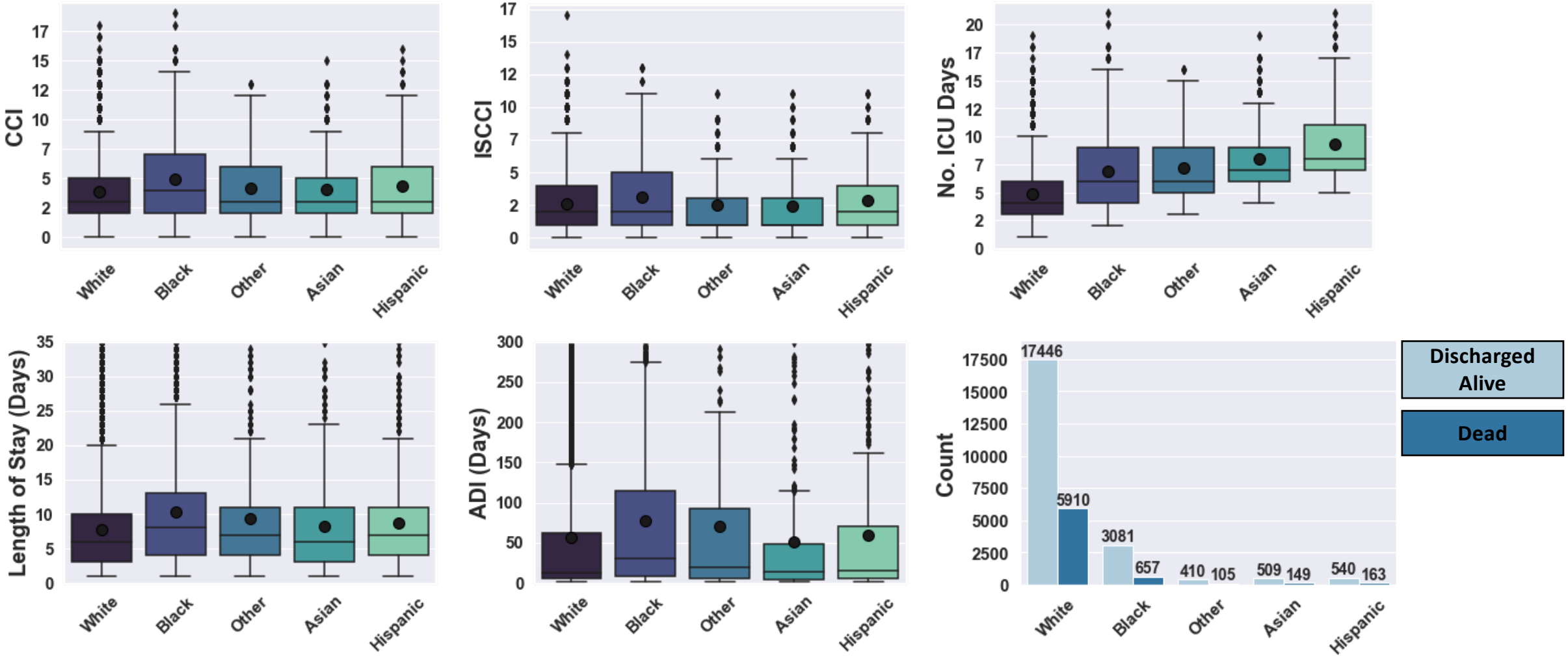
- Patient characteristics traditionally considered in procedural assessment were selected

- Basic features:
  - Source of Admission
  - Type of Admission
  - No. ICU Days
  - Length of Hospital Stay
  - Admission to death interval (ADI)
  - Discharge status [In-Hospital (IH) Mortality, Post-Discharge (PD) Mortality]
    - Note: IH Mortality includes discharge to acute end of life care (based on lit. definitions)

- Transformed features:
  - Charlson Comorbidity Index (CCI)
  - Ischemic Stroke CII (ISCCI)

Note: a program was deployed that performs web scraping to automatically transform diagnostic codes to CCI and ISCCI

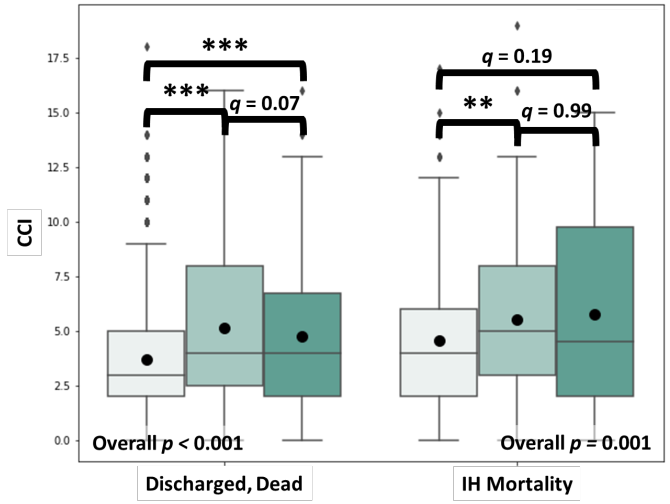
Clinical Measures and Outcomes by Demographic Group

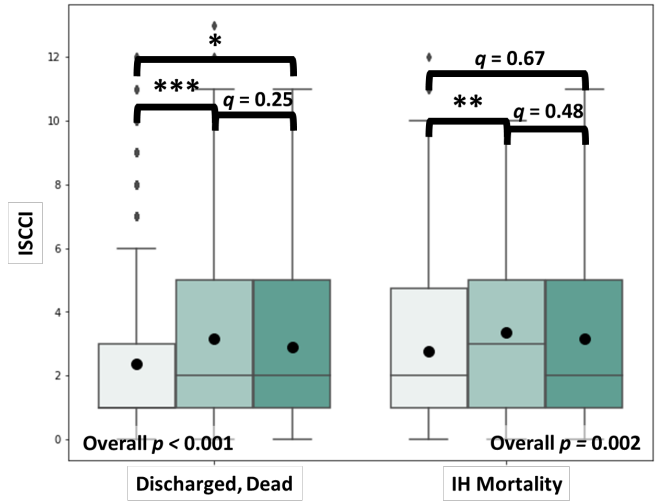


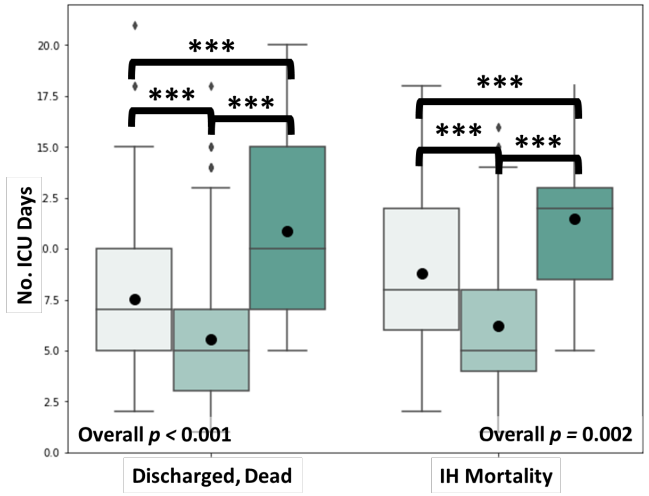
Takeaways: Race subpopulations demonstrate variance for selected features.

Statistical Analysis

Aim: Determine which clinical measures and outcomes are significantly different between races.





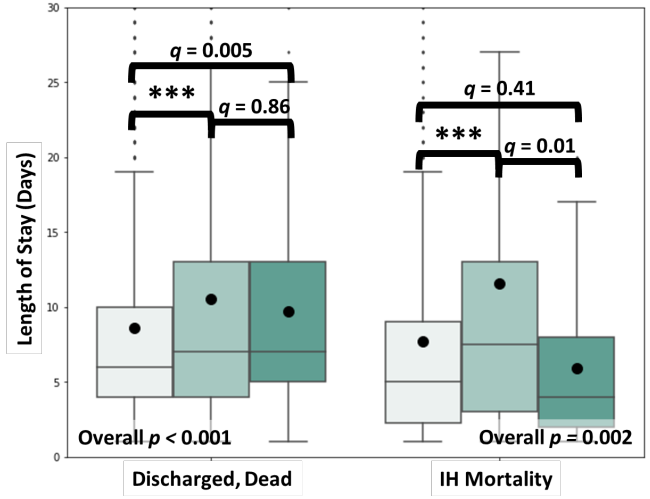


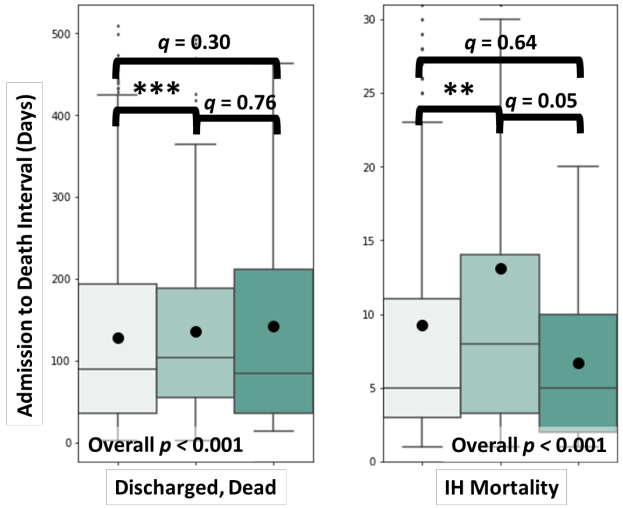
White

Black

Hispanic

$q < 0.05$ *
$q < 0.005$ **
$q < 0.001$ ***





Approach:

Norm: Anderson Darling

Comparison 3+: Kruskal-Wallis

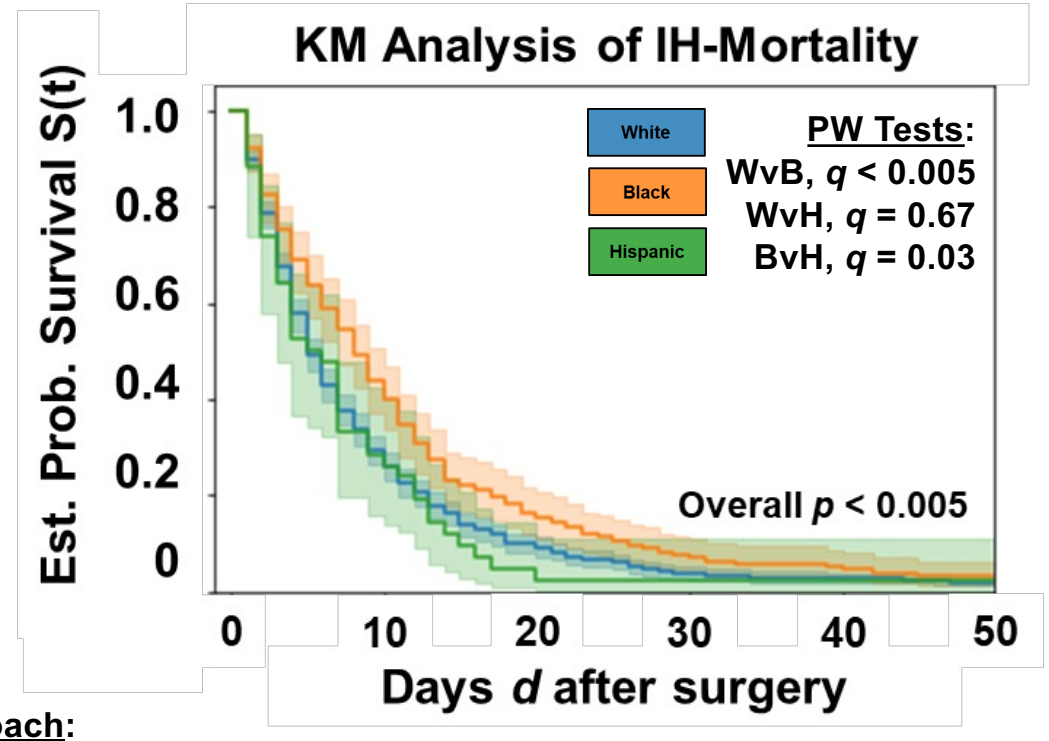
Multi-Comparison post-hoc Test: Dunn's

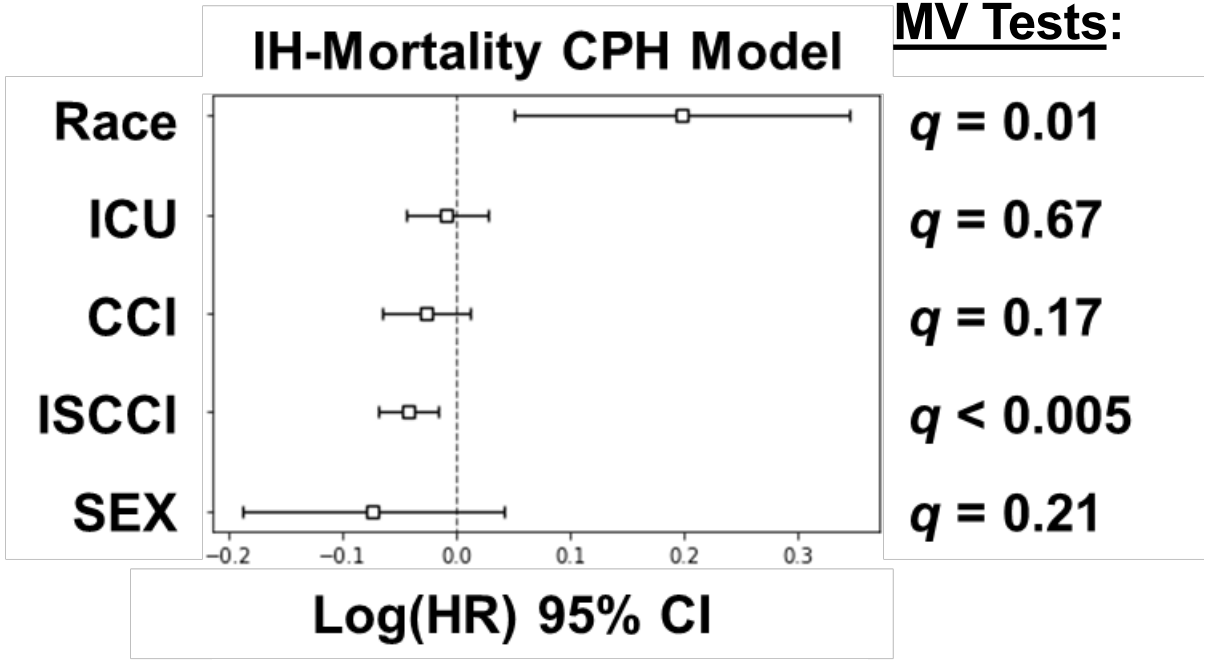
Correction for multiple testing: Benjamini-Hochberg

Statistical conclusions: W&B populations are always significantly different; W&H for a subset. No. ICU Days is the most discriminative feature.

Survival Analysis: IH-Mortality

Aim: Investigate if/what races are at an elevated risk of IH-mortality.





MV Tests:

$q = 0.01$

$q = 0.67$

$q = 0.17$

$q < 0.005$

$q = 0.21$

**Race: W&H=1/B=0, Sex: M=1/F=0*

Approach:

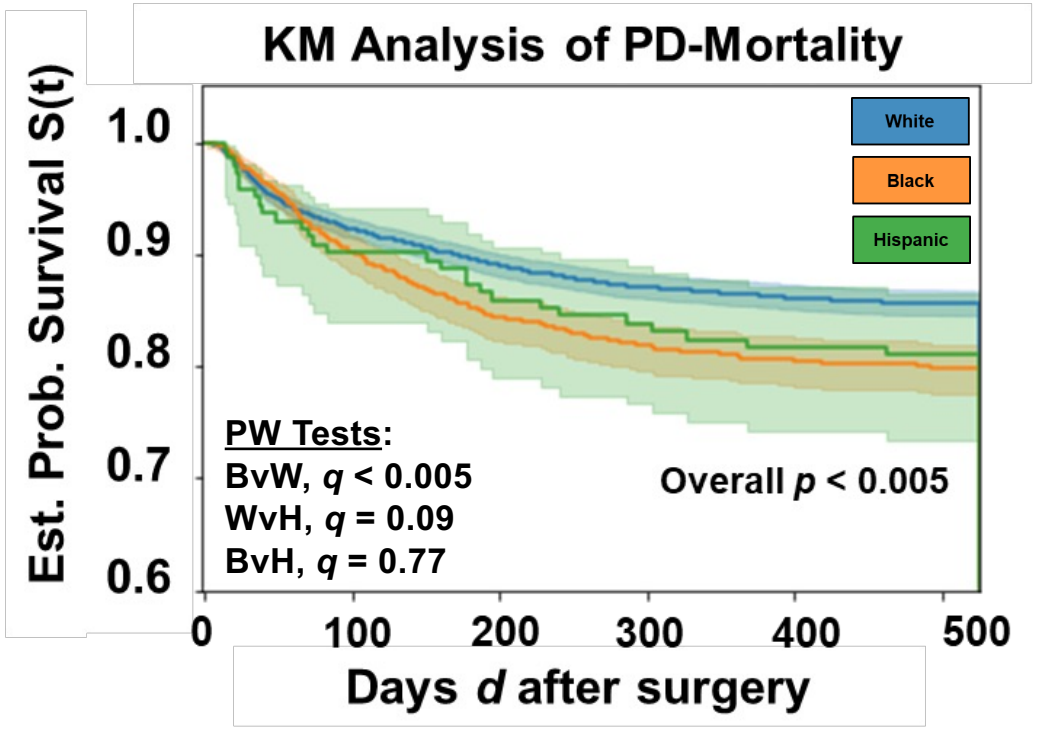
Kaplan-Meier (KM) Analysis → Pairwise (PW) Log-Rank Tests ****All with correction for multiple testing***

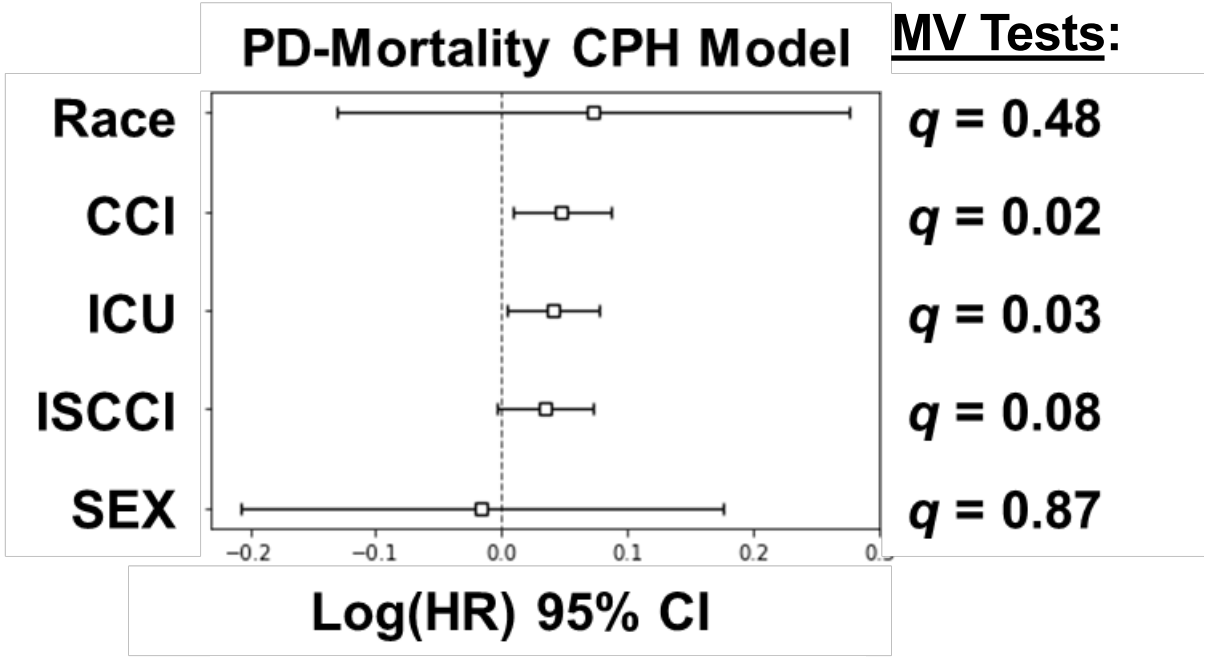
Cox Proportional Hazard (CPH) Model & Ratios → Multivariate Log-Rank Tests

Conclusions: Black populations have better short-term thrombectomy outcomes. Race is the leading hazard for mortality, with White and Hispanic populations at an elevated risk compared to black populations.

Survival Analysis: PD-Mortality

Aim: Investigate if/what races are at an elevated risk of IH-mortality.





MV Tests:

$q = 0.48$

$q = 0.02$

$q = 0.03$

$q = 0.08$

$q = 0.87$

**Race: B&H=1/W=0, Sex: M=1/F=0*

Conclusions: A sharp decline in survival is observed for black patients post-discharge. White populations have better long-term thrombectomy outcomes. Race is the leading hazard, although large variance suggests statistical insignificance. No. of ICU days and patient CCI are also strong and significant hazards.