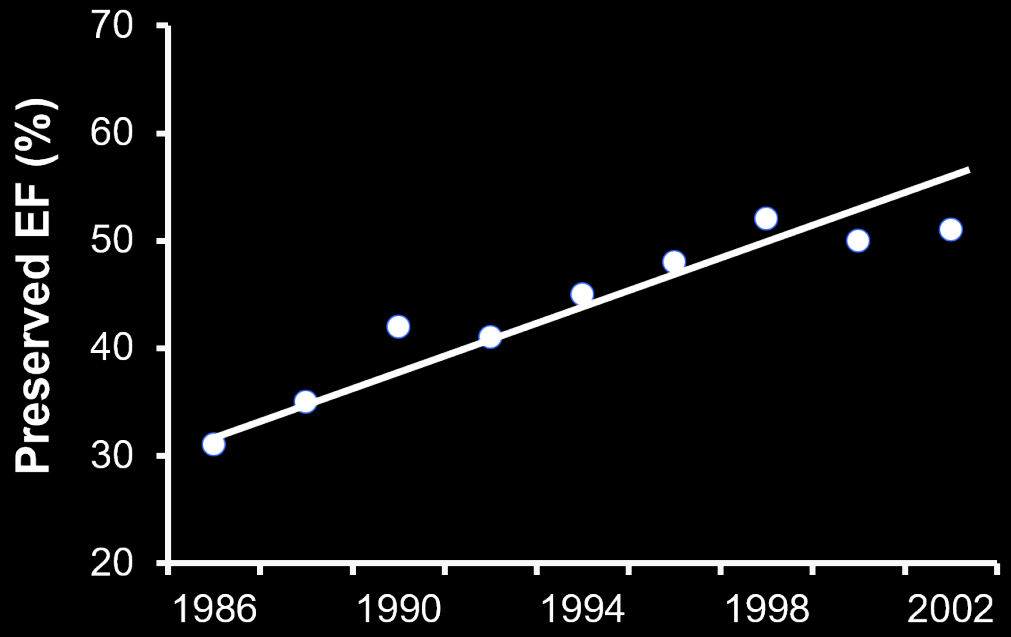
Phenomapping of Patients With Heart Failure With Preserved Ejection Fraction Using Machine Learning-Based Unsupervised Cluster Analysis

Matthew W. Segar, MD, MS
PGY-2, University of Texas Southwestern Medical Center

Disclosures: None

HFpEF: The New Face of HF

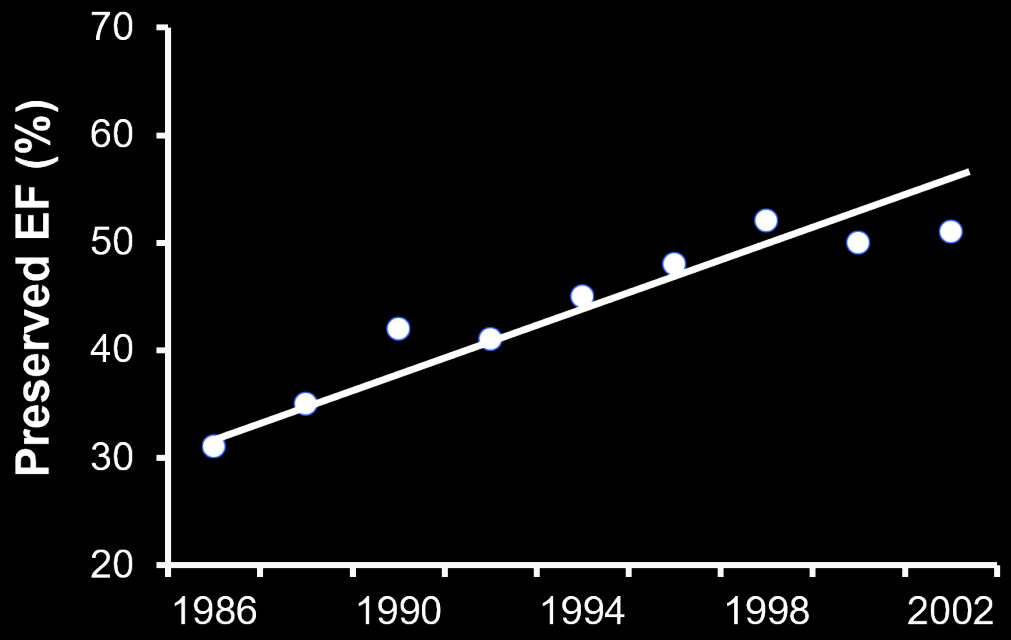
Increasing Prevalence



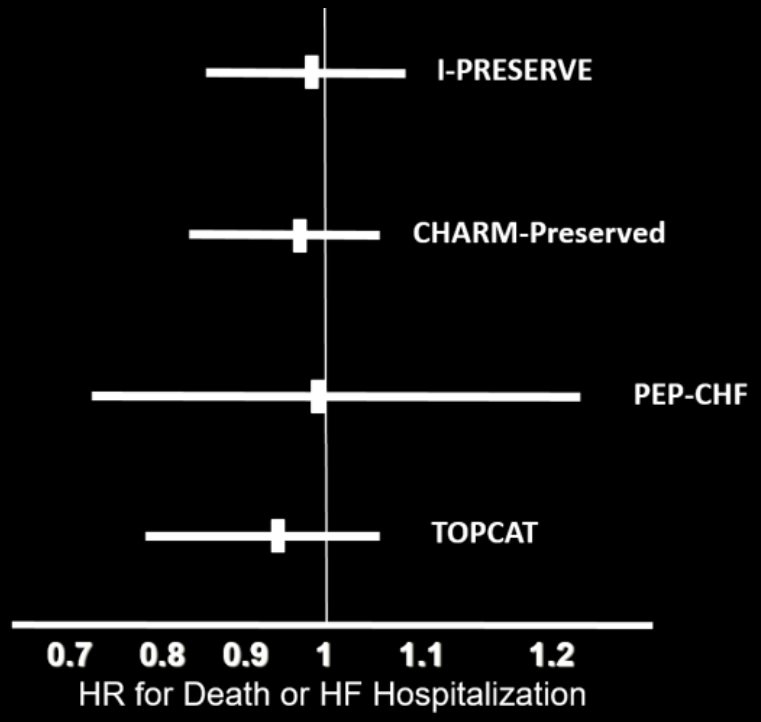
Owan, et al *NEJM* 2006

HFpEF: The New Face of HF

Increasing Prevalence

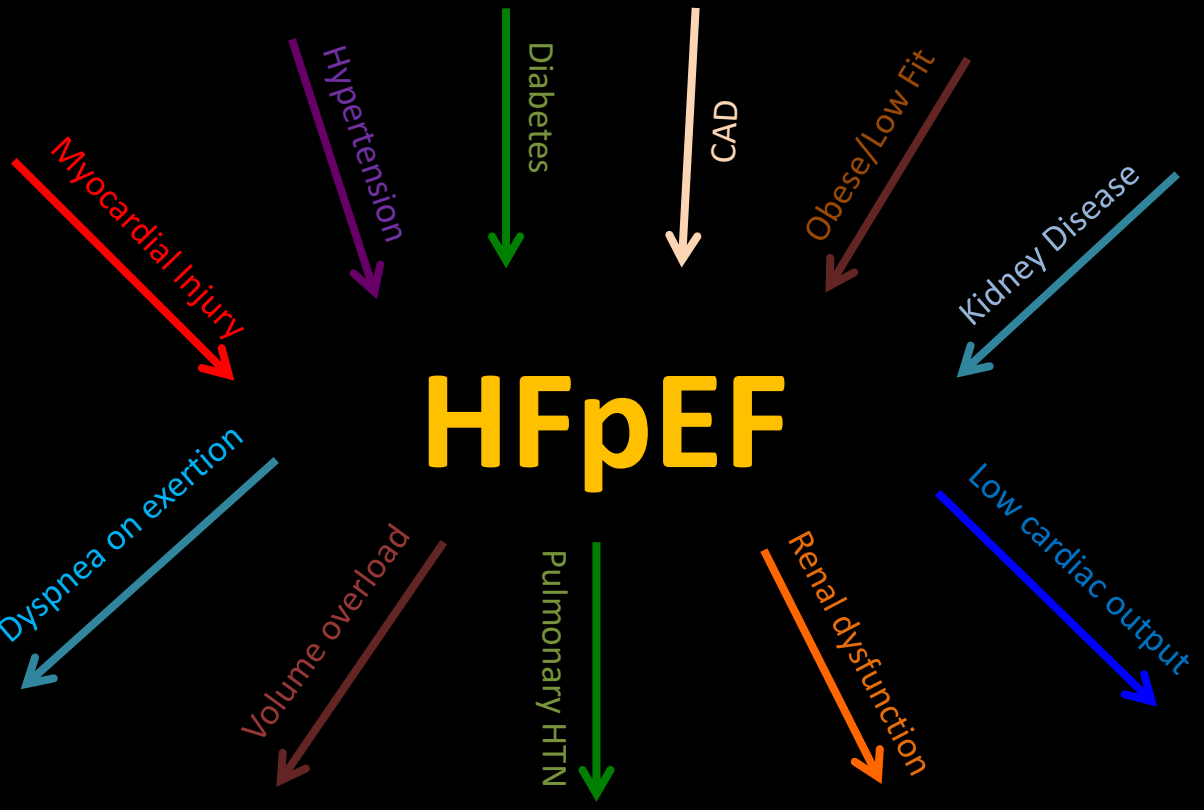


Refractory to Therapy



Owan, et al *NEJM* 2006

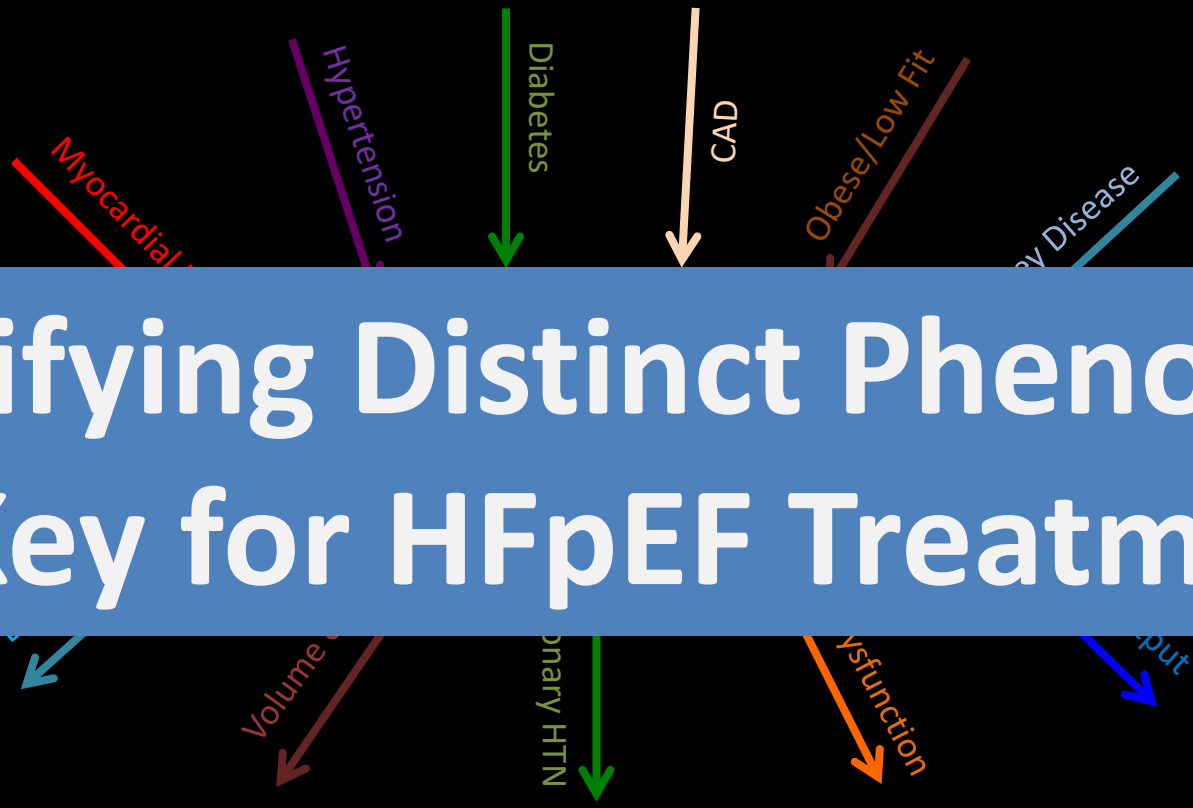
Different risk factors combinations leads to abnormal myocardial structure/function



Heterogenous clinical presentations

Shah SJ, et al. *Heart Fail Clin* 2014

Different risk factors combinations leads to abnormal myocardial structure/function



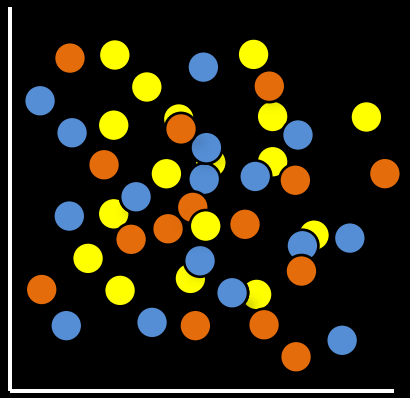
Identifying Distinct Phenotypes is Key for HFpEF Treatment

Heterogenous clinical presentations

Shah SJ, et al. *Heart Fail Clin* 2014

Central Hypothesis

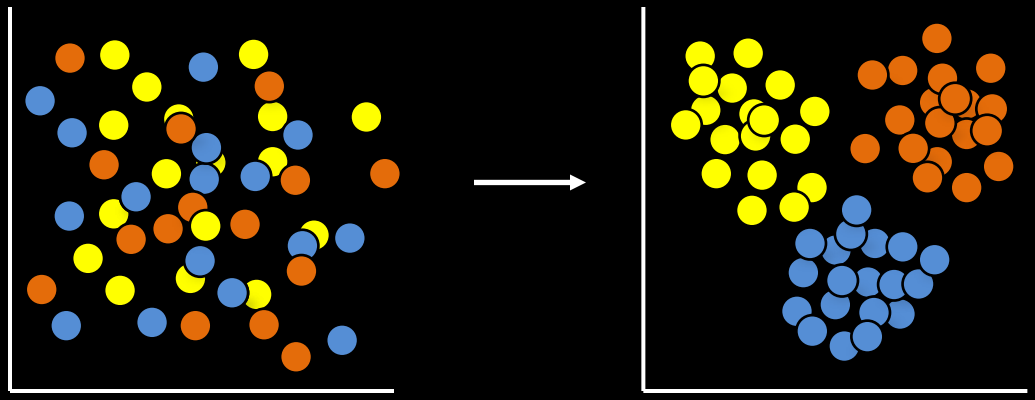
HFpEF Patients



Central Hypothesis

HFpEF Patients

Aim 1:
Determine Phenotypic Clusters

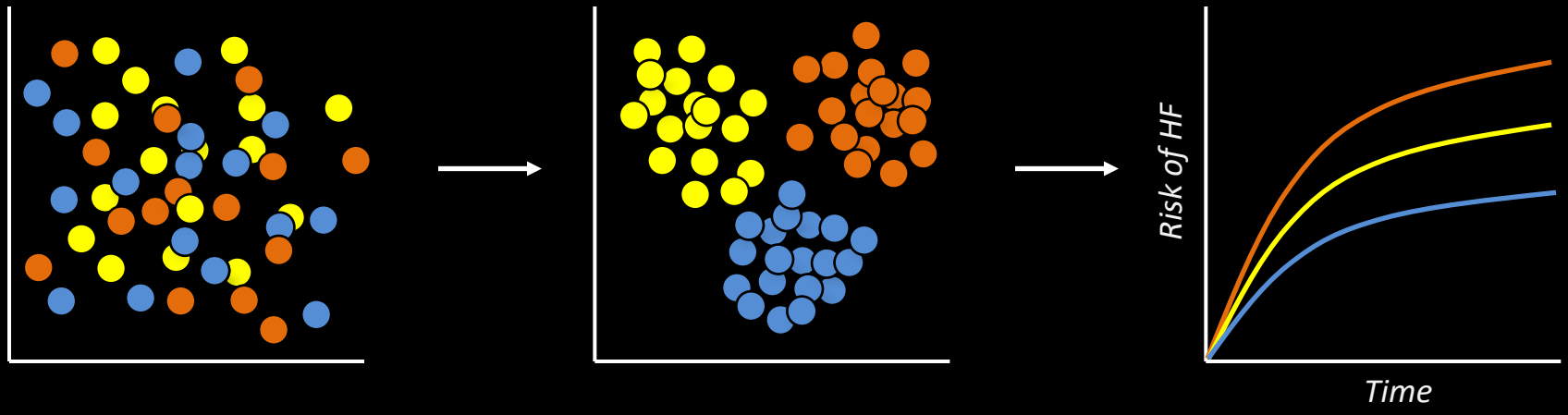


Central Hypothesis

HFpEF Patients

Aim 1:
Determine Phenotypic Clusters

Aim 2:
Unique Clinical Risk Profiles

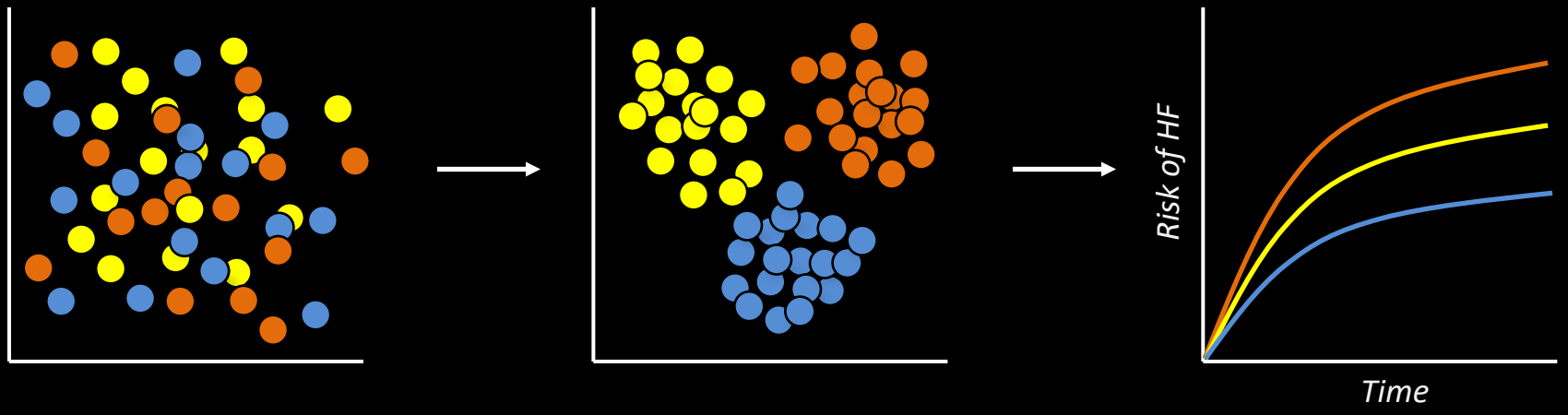


Central Hypothesis

HFpEF Patients

Aim 1:
Determine Phenotypic Clusters

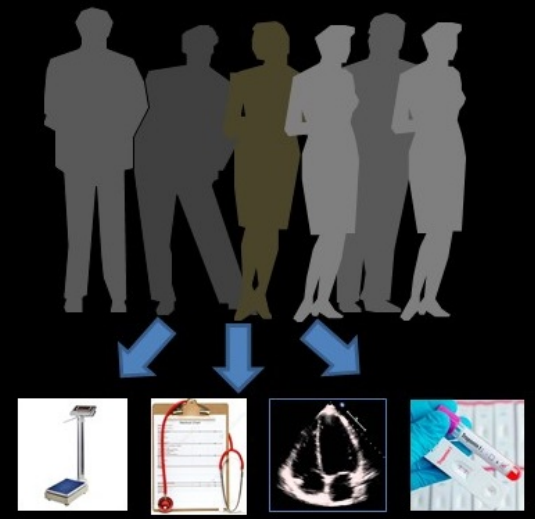
Aim 2:
Unique Clinical Risk Profiles



Hypothesis: Machine learning-based unsupervised clustering can identify distinct phenotypic subgroups in a mixed-data cohort of individuals with HFpEF.

Study Methodology

TOPCAT Trial
Stable HFpEF Patients (Americas)
(Spironolactone vs. Placebo)



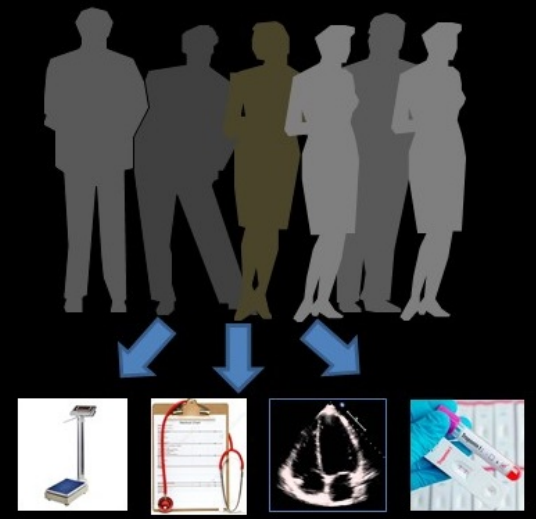
Body size | Clinical parameters | Echo | Labs

61 non-collinear variables
<10% missingness

N = 654

Study Methodology

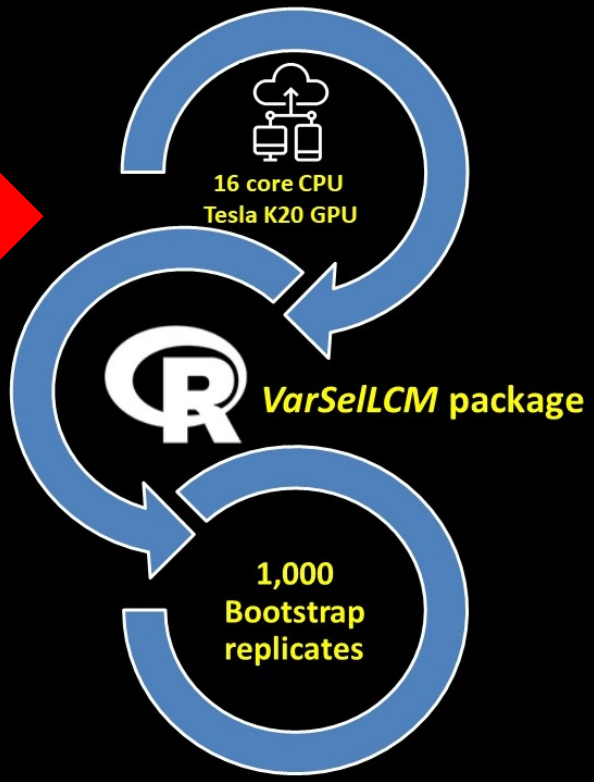
TOPCAT Trial
Stable HFpEF Patients (Americas)
(Spironolactone vs. Placebo)



Body size | Clinical parameters | Echo | Labs

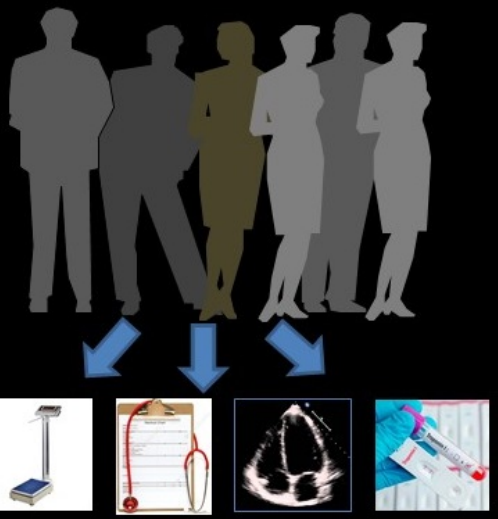
61 non-collinear variables
<10% missingness
N = 654

Machine Learning
Cluster Analysis



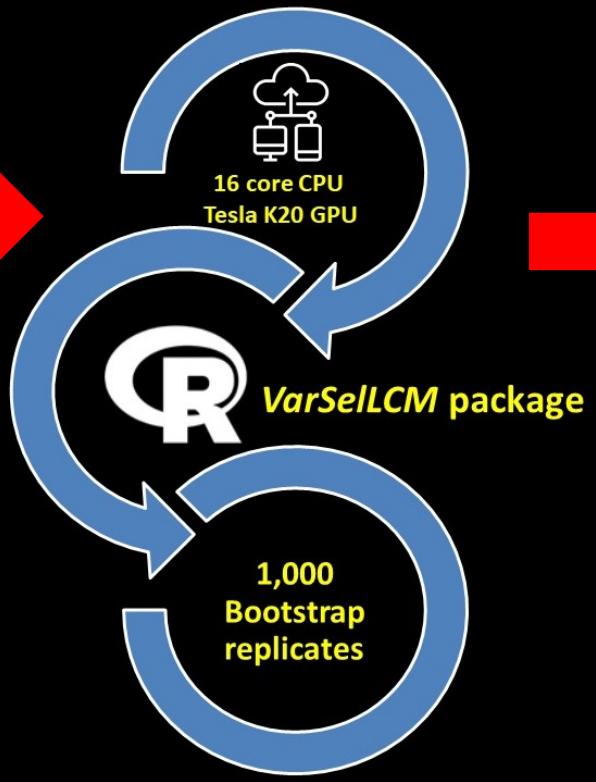
Study Methodology

TOPCAT Trial
Stable HFpEF Patients (Americas)
(Spironolactone vs. Placebo)



61 non-collinear variables
<10% missingness
N = 654

Machine Learning Cluster Analysis



Distinct Phenogroups

- **Phenogroup 1**
- **Phenogroup 2**
- **Phenogroup 3**

Clinical Characteristics of the Phenogroups

Phenogroup 1 (N = 136)

- Highest burden of CV risk factors:
  - Anemia
  - BMI
  - T2DM
  - eGFR
  - ASCVD
  - BNP/NT-ProBNP
- Echocardiographic:
  - Highest LV mass

Clinical Characteristics of the Phenogroups

Phenogroup 1 (N = 136)

- Highest burden of CV risk factors:
  - Anemia
  - BMI
  - T2DM
  - eGFR
  - ASCVD
  - BNP/NT-ProBNP
- Echocardiographic:
  - Highest LV mass

Phenogroup 2 (N = 108)

- Lowest burden of:
  - T2DM
  - Dyslipidemia
  - ASCVD
- Echocardiographic:
  - Highest diastolic dysfunction

Clinical Characteristics of the Phenogroups

Phenogroup 1 (N = 136)

- Highest burden of CV risk factors:
  - Anemia
  - BMI
  - T2DM
  - eGFR
  - ASCVD
  - BNP/NT-ProBNP
- Echocardiographic:
  - Highest LV mass

Phenogroup 2 (N = 108)

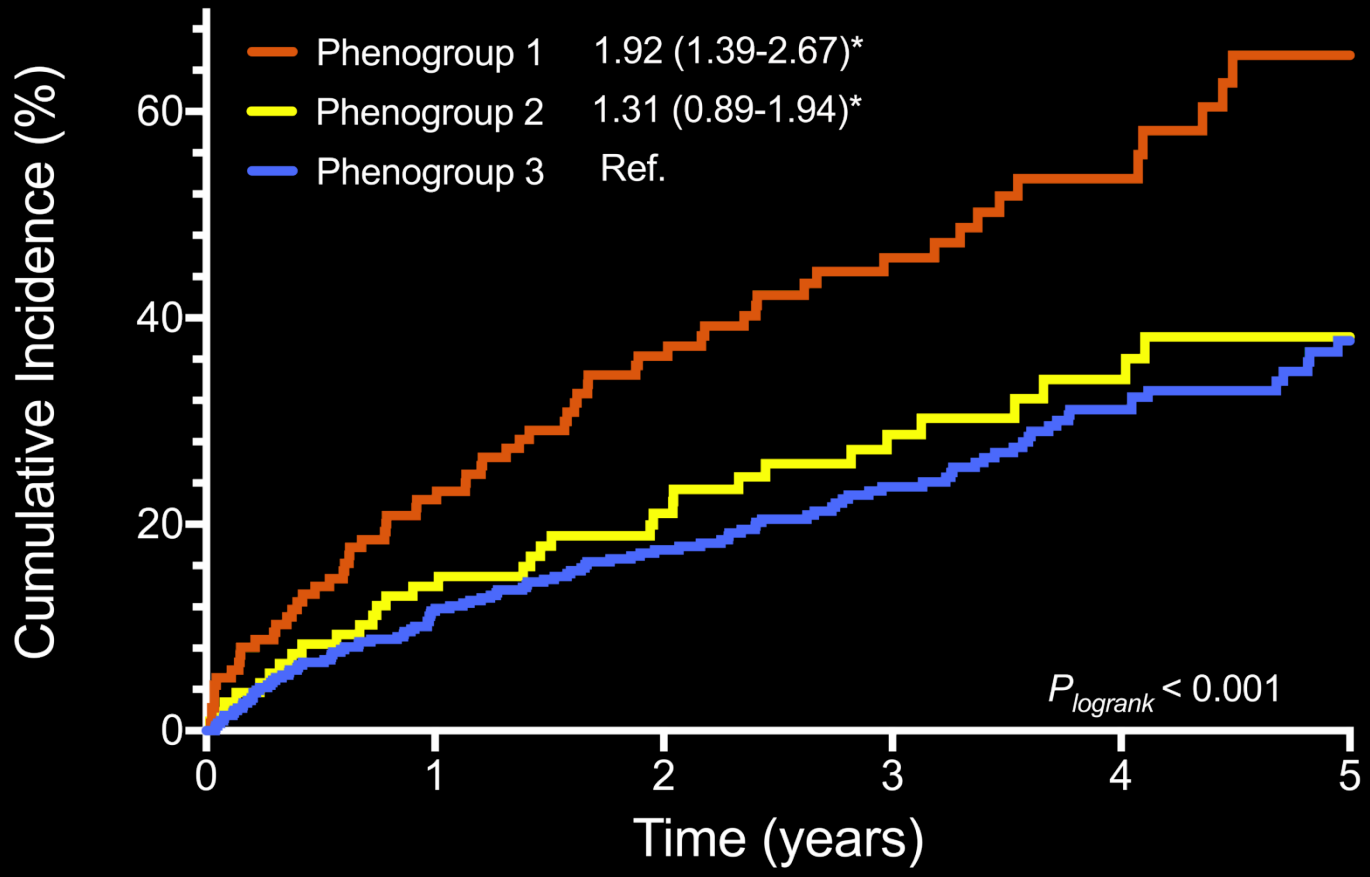
- Lowest burden of:
  - T2DM
  - Dyslipidemia
  - ASCVD
- Echocardiographic:
  - Highest diastolic dysfunction

Phenogroup 3 (N = 410)

- Lowest burden of:
  - Obesity
  - BNP/NT-ProBNP
- Echocardiographic:
  - Lowest diastolic dysfunction
  - Lower LA size
  - Lower PA pressure

Clinical Validity of the Phenomapping Approach

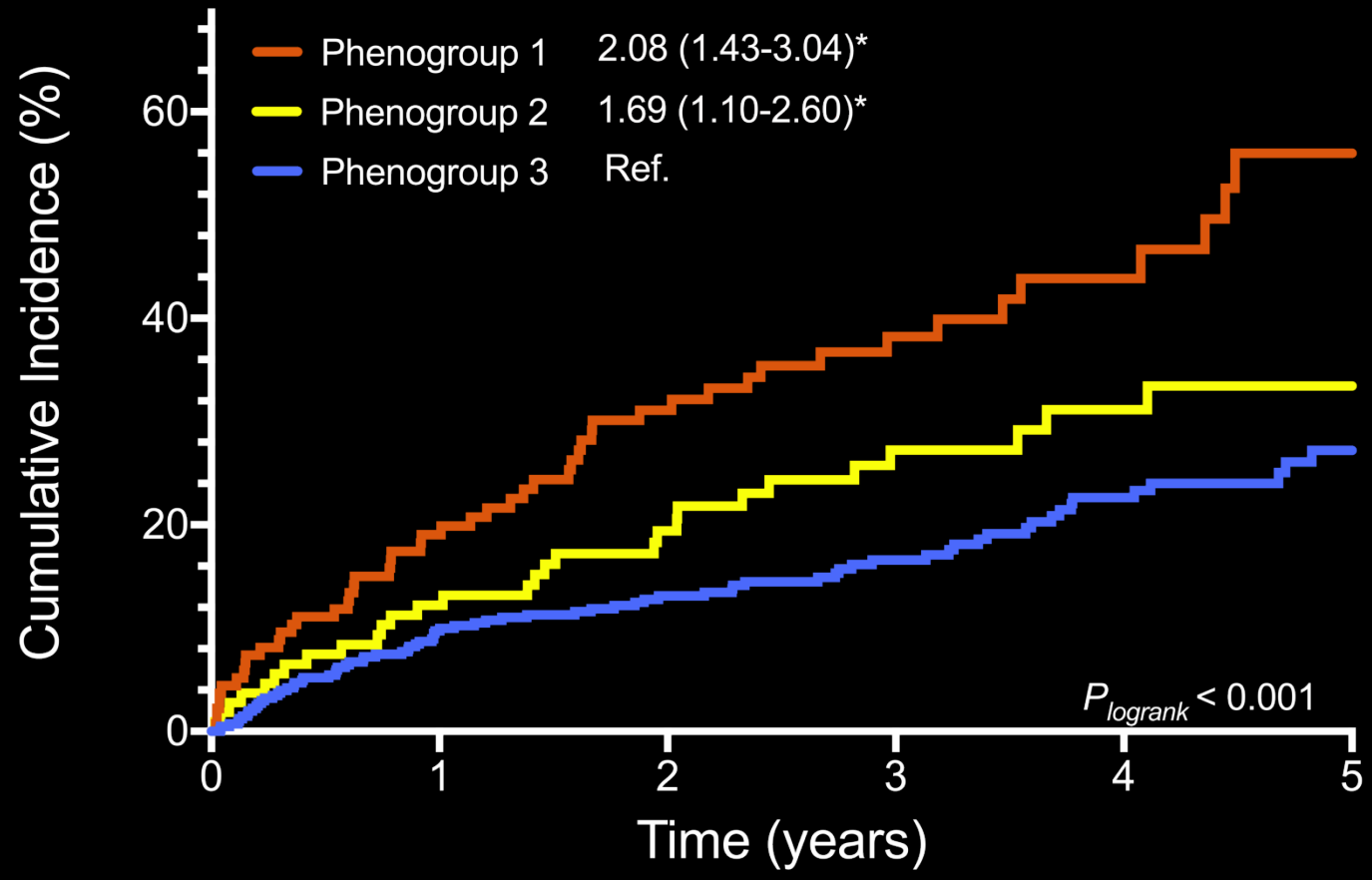
Primary composite endpoints[#] (N = 209)



**Hazard ratio for outcome adjusted for age, diabetes status, randomization strata, and treatment arm*
#Composite of aborted cardiac arrest, HF hospitalization, or CV death.

Clinical Validity of the Phenomapping Approach

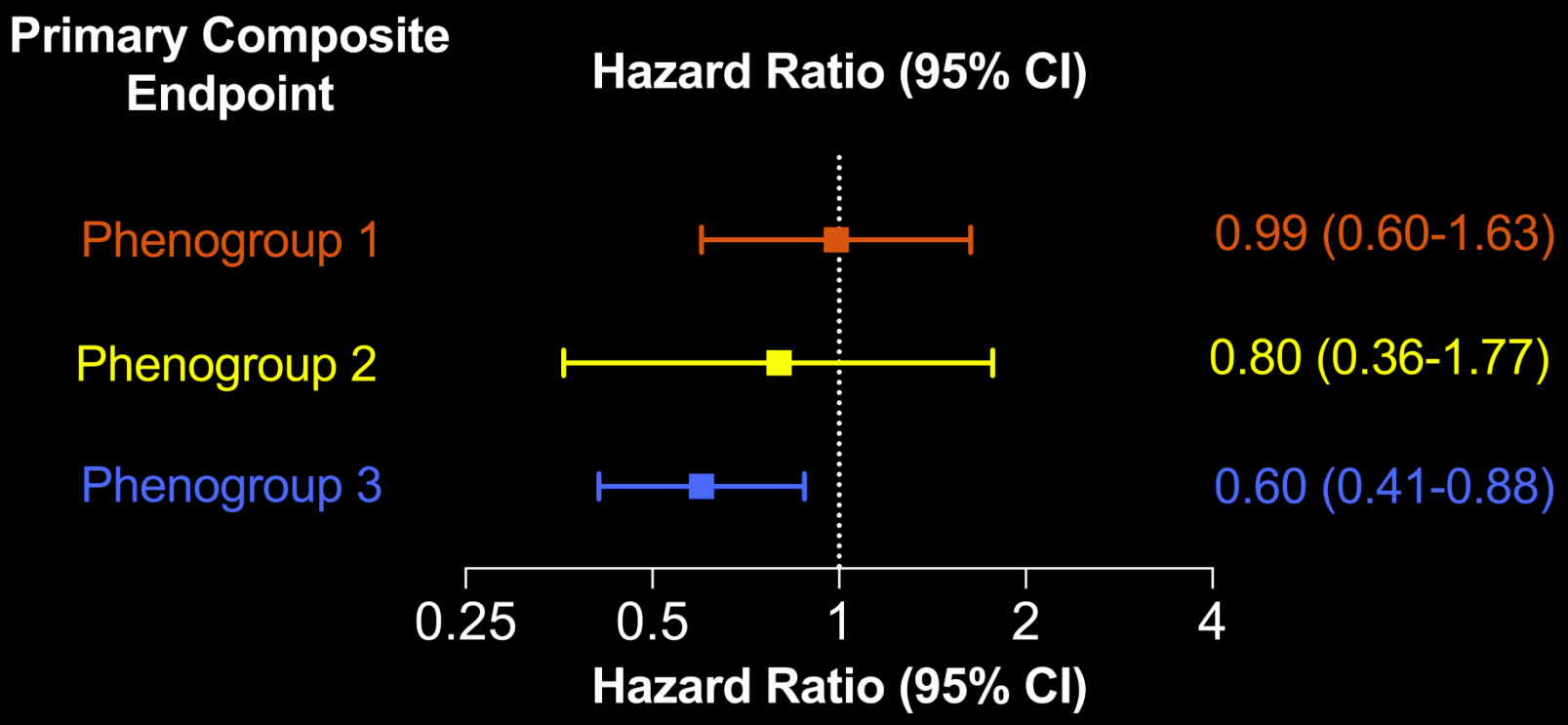
HF hospitalization (N = 157)



**Hazard ratio for outcome adjusted for age, diabetes status, randomization strata, and treatment arm*

Efficacy of Spironolactone across Phenogroups

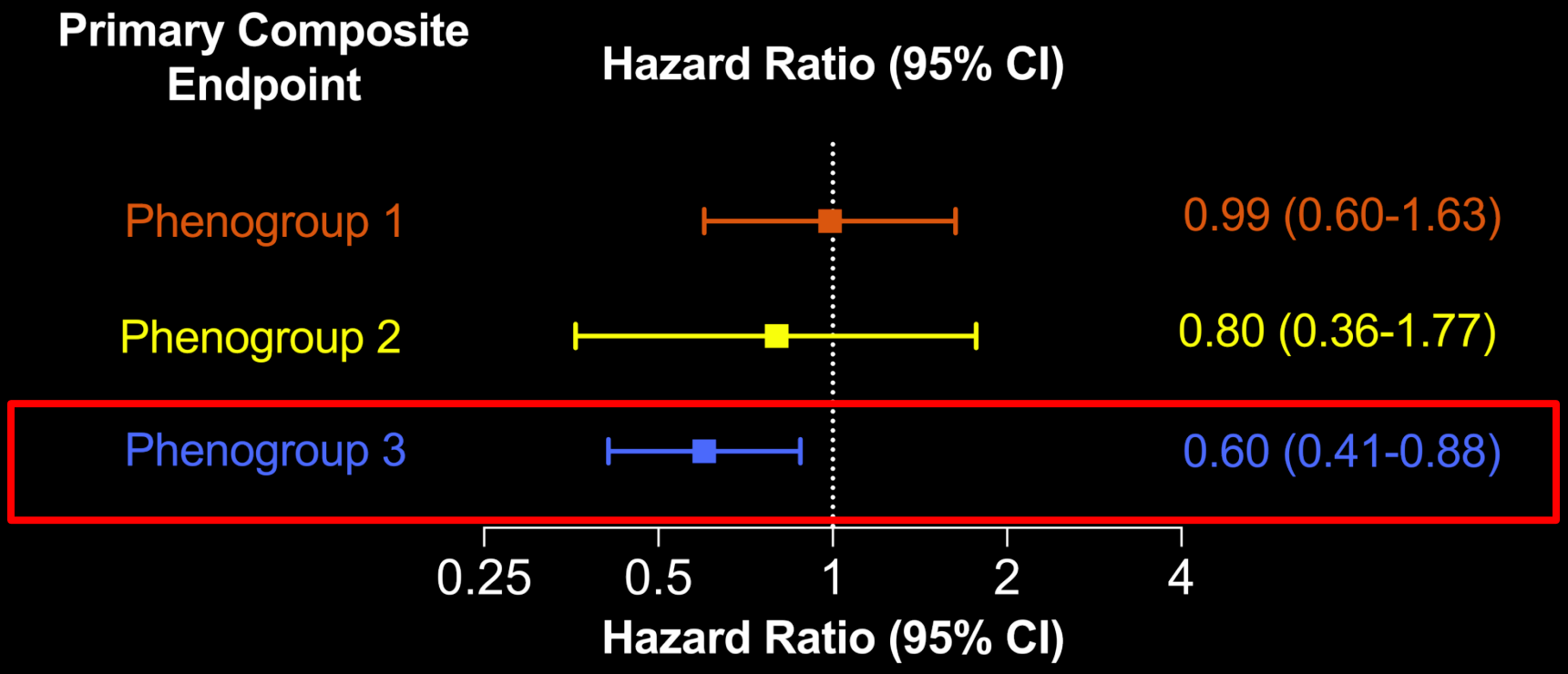
P-interaction (Treatment arm X Phenogroup): 0.10



Hazard ratio for outcome adjusted for age, diabetes status, randomization strata, and treatment arm
Composite of aborted cardiac arrest, HF hospitalization, or CV death.

Efficacy of Spironolactone across Phenogroups

P-interaction (Treatment arm X Phenogroup): 0.10



Hazard ratio for outcome adjusted for age, diabetes status, randomization strata, and treatment arm
Composite of aborted cardiac arrest, HF hospitalization, or CV death.

External Validation of the Phenomapping Approach

RELAX Trial Cohort (N = 198 stable HFpEF patients)

Outcome

- Composite rank score during follow up (6 months) based on significant clinical events
  - Death
  - Time to CV hospitalization
  - Worsening HF symptoms

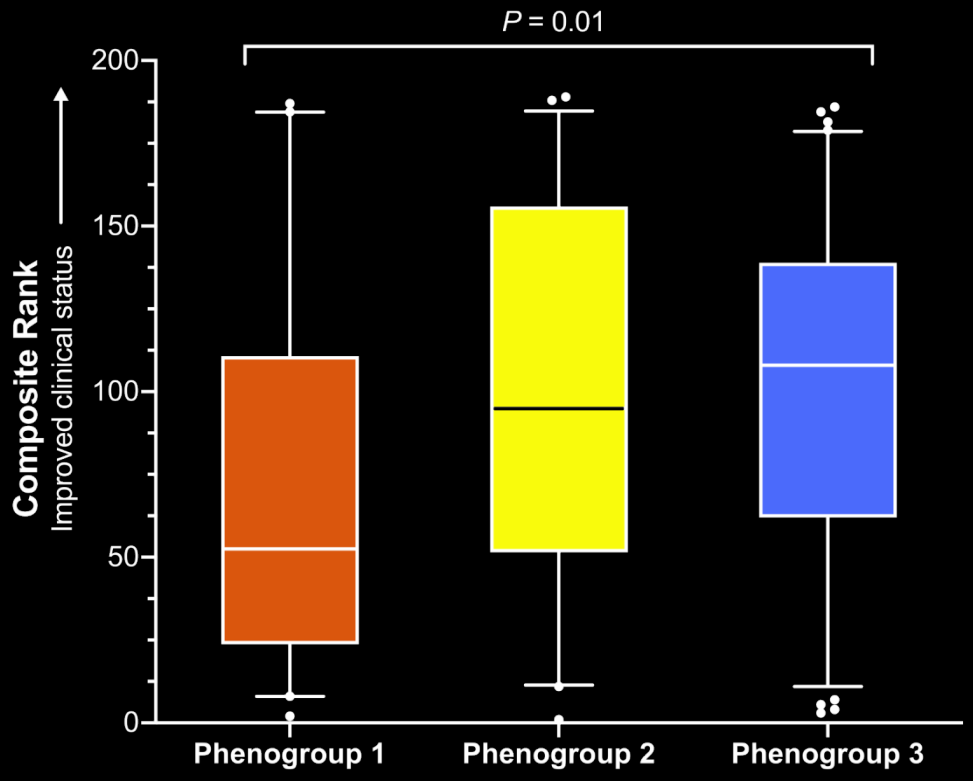
- Lower score indicates worse prognosis

External Validation of the Phenomapping Approach

RELAX Trial Cohort (N = 198 stable HFpEF patients)

Outcome

- Composite rank score during follow up (6 months) based on significant clinical events
  - Death
  - Time to CV hospitalization
  - Worsening HF symptoms
- Lower score indicates worse prognosis



Limitations

- Exploratory analysis due to the negative primary results of the TOPCAT trial
- Missingness in other relevant clinical biomarkers
- Need larger cohort to test treatment interaction

Conclusion

- Machine learning-based cluster analysis can identify phenogroups of patients with HFpEF with distinct clinical characteristics and long-term outcomes.

Acknowledgements

Primary Mentor

- **Ambarish Pandey, MD**

UTSW

- Jarrett Berry, MD
- Justin Grodin, MD, MPH
- Kershaw Patel, MD
- Mujeeb Basit, MD, MS
- Duwayne Willett, MD, MS

Computational/Resource Support

UT Southwestern Medical Center
Lyda Hill Department of Bioinformatics
BioHPC

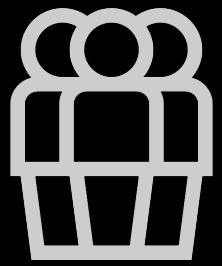
Research Funding Support

Southwestern Health Resources
Texas Health Resources®
UT Southwestern Medical Center

Clinical Characteristics of the Phenogroups

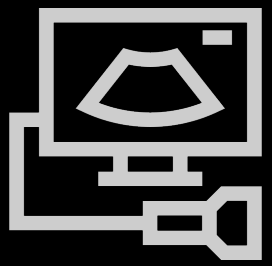
Baseline Characteristics	Phenogroup 1 (N =136)	Phenogroup 2 (N = 108)	Phenogroup 3 (N =410)
BMI, kg/m2	35	33	31
NYHA Class III or above (%)	52	28	33
Diabetes (%)	74	26	44
Prior PCI (%)	22	10	19
Peripheral Arterial disease (%)	16	6	10
Dyslipidemia (%)	78	66	67
Glucose, mg/dl	138	109	123
Total Bilirubin, mg/dl	0.67	0.83	0.63
Estimated GFR, ml/min	52	66	67
NT-ProBNP, pg/ml	2913	1470	1419
Peak TR velocity, cm/s	311	292	266
LV mass, mg	236	210	222
Severe Diastolic dysfunction (%)	28	33	1.7

Methods



TOPCAT

- 3445 participants
- 156 variables

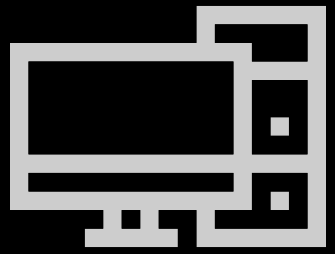


Echocardiographic

- 935 participants
- 156 variables

Americas

- 654 participants
- 82 variables



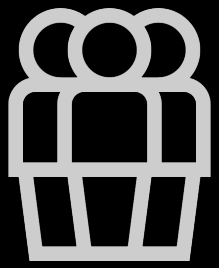
Final Analysis

- 654 participants
- 61 variables

Pitt, *et al.* "Spironolactone for HFpEF" *NEJM* (2014)

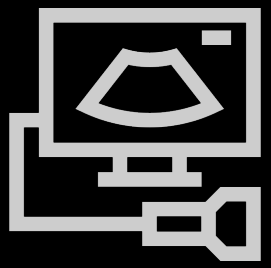
Pfeffer, *et al.* "Regional variation in TOPCAT" *Circulation* (2015)

External Validation



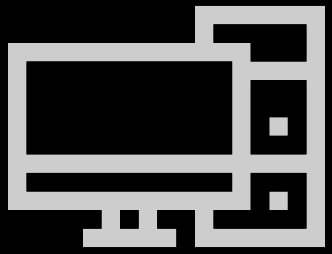
RELAX

- 216 participants
- 38 variables



Echocardiographic

- 214 participants
- 38 variables

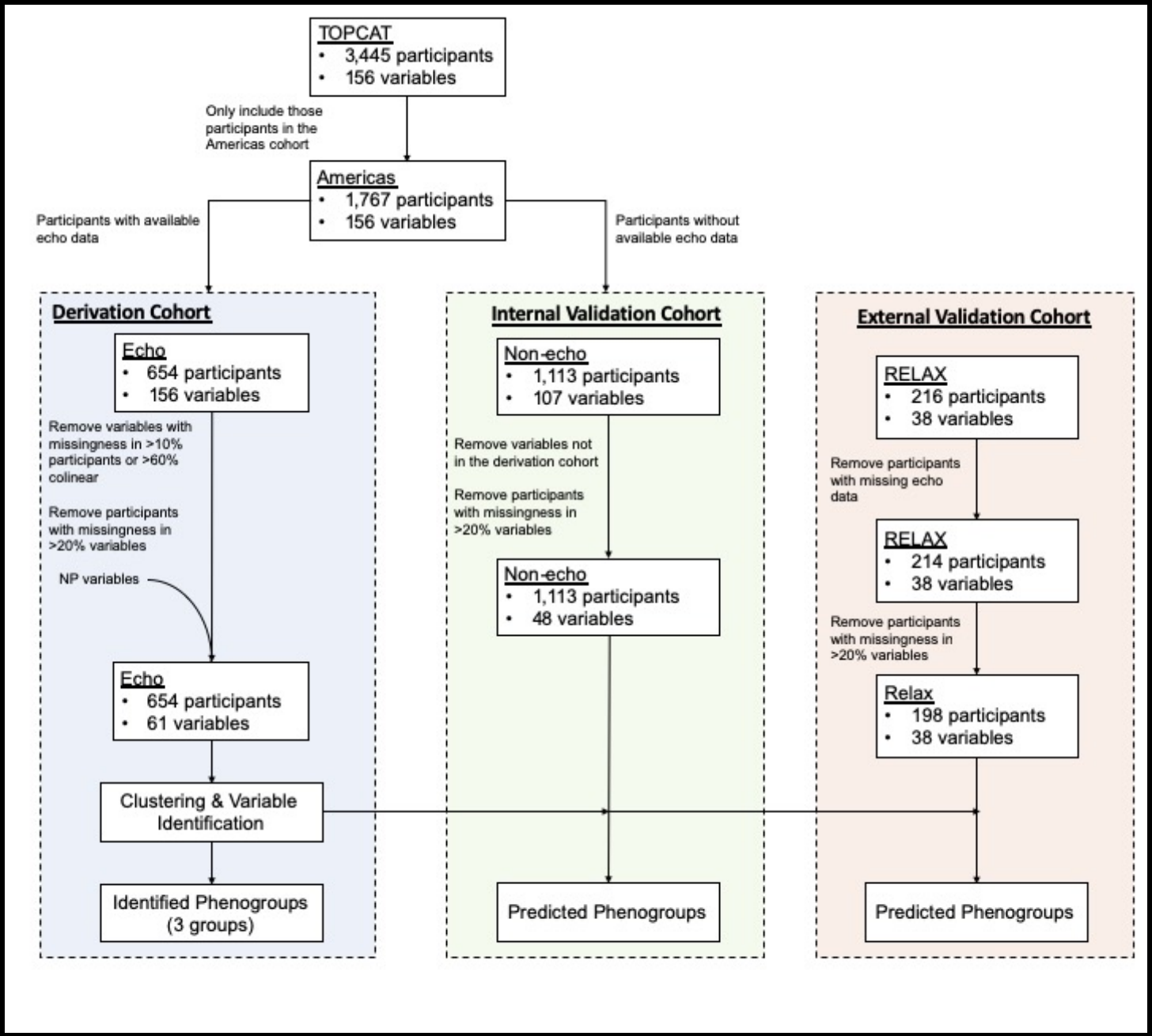


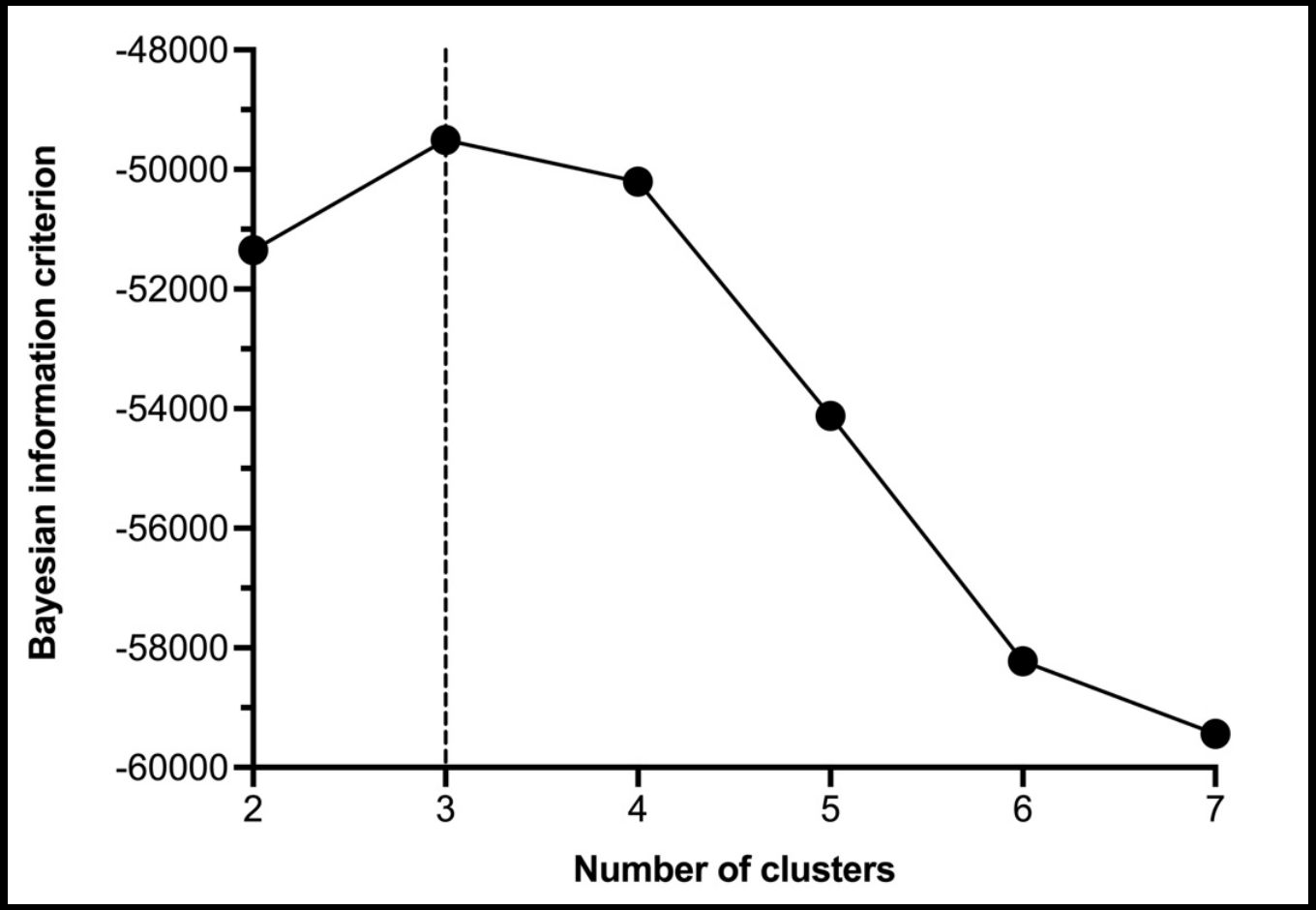
Final Analysis

- 198 participants
- 38 variables

Redfield, *et al.* "PDE-5 in HFpEF" *JAMA* (2013)

	TOPCAT (n=654)	RELAX (n=198)	P-value
Age, years	71.18 (9.85)	68.58 (10.20)	0.001
Male, n	339 (51.8)	99 (50.0)	0.710
History of MI, n	138 (21.1)	20 (10.1)	0.001
History of COPD, n	109 (16.7)	35 (17.7)	0.829
History of HTN, n	595 (91.1)	90 (45.5)	<0.001
History of PAD, n	70 (10.7)	29 (14.6)	0.167
History of ICD, n	14 (2.1)	1 (0.5)	0.22
Diabetes category, n			0.192
No diabetes	345 (52.8)	112 (56.6)	
Insulin dependent	159 (24.3)	36 (18.2)	
Non-insulin dependent	149 (22.8)	50 (25.3)	
Current smoker, n	41 (6.3)	25 (12.7)	0.005
Ever smoker, n	332 (54.2)	124 (62.9)	0.04
NYHA Class (III or above),	236 (36.3)	105 (53.0)	<0.001
HR, bpm	69.45 (11.64)	69.83 (11.73)	0.682
SBP, mmHg	126.89 (15.77)	127.88 (17.43)	0.451
BMI, kg/m^2	33.80 (7.99)	33.94 (7.02)	0.823
NA, mg/dL	139.47 (3.04)	139.68 (3.23)	0.405
K, mg/dL	4.15 (0.43)	4.32 (0.50)	<0.001
CO2, mg/dL	28.49 (3.63)	27.44 (3.22)	0.002
BUN, mg/dL	25.13 (12.29)	27.66 (14.18)	0.021
Glucose, mg/dL	123.95 (61.23)	123.45 (50.37)	0.923
BNP, pg/mL	364.92 (265.7)	413.0 (455.1)	0.410
Hb, U/L	12.71 (1.64)	12.90 (1.49)	0.135
AST, U/L	25.87 (14.35)	29.07 (14.44)	0.007
Total bilirubin, U/L	0.67 (0.45)	0.63 (0.34)	0.269
Albumin, U/L	3.84 (0.44)	3.98 (0.52)	0.002
QRS duration, ms	106.53 (31.09)	106.85 (33.65)	0.902
ACE inhibitor, n	343 (52.4)	88 (44.4)	0.058
ARB, n	189 (28.9)	55 (27.8)	0.829
Beta blocker, n	530 (81.0)	147 (74.2)	0.048
Calcium channel blocker, n	252 (38.5)	59 (29.8)	0.031
Diuretic, n	586 (89.6)	171 (86.4)	0.254
Aspirin, n	372 (56.9)	122 (61.6)	0.271
Statin, n	443 (67.7)	127 (64.1)	0.392
Warfarin, n	229 (35.0)	83 (41.9)	0.092
SSRI, n	93 (14.2)	61 (30.8)	<0.001
Ejection fraction, %	59.58 (7.80)	60.81 (6.52)	0.044
LV diastolic diameter, cm	4.80 (0.57)	4.67 (0.65)	0.015
E/a ratio	1.35 (0.73)	1.87 (1.36)	<0.001
TR velocity, cm	281.31 (46.38)	287.69 (43.25)	0.163
LA area, cm^2	20.20 (5.63)	22.29 (7.06)	0.011
H2FPEF score	6.04 (1.94)	6.47 (1.21)	0.100


TOPCAT
• 3,445 participants
• 156 variables
Only include those participants in the Americas cohort
Americas
• 1,767 participants
• 156 variables
Participants with available echo data
Participants without available echo data
Derivation Cohort
Echo
• 654 participants
• 156 variables
Remove variables with missingness in >10% participants or >60% colinear
Remove participants with missingness in >20% variables
NP variables
Echo
• 654 participants
• 61 variables
Clustering & Variable Identification
Identified Phenogroups (3 groups)
Internal Validation Cohort
Non-echo
• 1,113 participants
• 107 variables
Remove variables not in the derivation cohort
Remove participants with missingness in >20% variables
Non-echo
• 1,113 participants
• 48 variables
Predicted Phenogroups
External Validation Cohort
RELAX
• 216 participants
• 38 variables
Remove participants with missing echo data
RELAX
• 214 participants
• 38 variables
Remove participants with missingness in >20% variables
Relax
• 198 participants
• 38 variables
Predicted Phenogroups


-48000
-50000
-52000
-54000
-56000
-58000
-60000
Bayesian information criterion
2
3
4
5
6
7
Number of clusters