Computational Methods to Improve Clinical Decision Science Related to Pulmonary Arterial Hypertension

Submitted in partial fulfillment of the requirements for the

degree of

Doctor of Philosophy in

Biomedical Engineering

Jacqueline Victoria Scott

B.S., Biomedical Engineering, University of Michigan, Ann Arbor
M.S., Physiology, University of Michigan, Ann Arbor

Carnegie Mellon University
Pittsburgh, PA

May 2022

Acknowledgements

First and foremost, I would like to thank my PhD advisor, Dr. Keith Cook, for his guidance, mentorship, and support. Even prior to my embarkment into a doctoral degree, Dr. Cook was a research mentor and role model for me, coaching me through several career and life stages. He has been a fierce advocate of my success in all aspects of life and has acted as a guiding light for almost a decade. I cannot imagine achieving the great honor of a doctoral degree without Dr. Cook's exceptional guidance.

I would also like to thank my doctoral committee members, Dr. Newell Washburn, Dr. Patjanaporn Chalacheva, and Dr. Manreet Kanwar.

Next, I would like to thank my parents, Camelia Scott and Timothy Menifield, for raising me to always value education and the pursuit of knowledge. I would also like to thank my sister, Stephanie Scott, who has loved me unwaveringly throughout my entire life.

Frankly, there are too many people I would like to thank amongst my family, friends, and colleagues to list them all. Squeezing in as many names as possible, my utmost gratitude goes to the following people and beyond: Lisa, Grace, and Bailey Tapert; Britney Menifield, Mackenzie Caple, Jay Satonik, Ariana Mirian, Catherine Salvatore, Kalliope Roberts, Megan Debari, Norman Stockbridge, Christine Garnett, Raymond Benza, Steven Kawut, Ethan Schonbrun, and many others. Thank you all for seeing me through one of the greatest achievements I've had the privilege to pursue.

Finally, this work would not be possible without the financial support of the National Institute of Health (R01 HL134673 and R01 HL089043), and the Oak Ridge Institute for Science and Education.

Abstract

Pulmonary arterial hypertension (PAH) remains a deadly and rare disorder of the pulmonary vasculature. A combination of vasoconstriction and cellular proliferation of the pulmonary arterial lumen results in increased mean pulmonary arterial pressure, straining the right heart and eventually causing heart failure. Despite the development of a wide range of pharmaceutical treatments for PAH, median survival of this condition remains a paltry seven years. Treatment guidance for PAH depends significantly on a clinician's ability to assess their patient's risk of mortality, but all risk assessment methods remain limited in their accuracy and usability.

This dissertation examines the ways in which risk assessment can be used to improve clinical decision science related to pulmonary arterial hypertension, then explores new modeling methodologies to improve upon current risk assessment standards. Chapter 1 discusses the background of challenges related to clinical decision science, specific to pulmonary arterial hypertension. Chapter 2 examines how risk assessment tools can be used to improve the efficiency of clinical trials for pharmaceutical treatments of PAH. Chapter 3 explores how improved risk stratification can reveal differences in treatment response between low and high-risk patients. Chapter 4 examines how novel machine learning methodologies can be employed to improve PAH risk assessment tools. Finally, Chapter 5 studies the potential use of a novel physiological model of right ventricular energetics as a means of improving clinical understanding of right heart failure.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Background

Need for Improvements in Clinical Decision Science

In 2000, the landmark publication *To Err is Human: Building a Safer Health System*, made two things abundantly clear: 1) even the best clinician is no better than the healthcare system they must function within, 2) bad systems lead to bad outcomes.(1) Although *To Err is Human* was published more than 20 years ago, the lessons from that report could not be more relevant. A recent meta-analysis determined that medical errors result in 22,165 preventable deaths per year.(2)

While medical errors are often thought of as obvious, egregious mistakes such as improper medication dosing or performing surgery on the wrong limb, the definition of a medical error allows for several shades of nuance. A medical error is deemed any preventable adverse effect of medical care.(2) A clinician may make what they believe is a reasonable decision for a patient that is ultimately poor because the severity of the patient's disease cannot be readily assessed. Further, in a clinical environment where time is of the essence, inconvenient clinical tests can contribute to poor decision-making. The cost of poor clinical decisions extends even to the design of clinical trials, as 50% of drugs fail their Phase 3 clinical trial, at enormous costs to pharmaceutical companies and the patients that depend on new treatment development.(3)

As the book *Improving Outcomes with Clinical Decision Support: An Implementer's Guide* astutely proposes, clinical decision support is all about "getting *the right information* to the *right person* at the *right point* in workflow in the *right intervention*

format through the *right channel*".(4) This dissertation specifically focuses on one aspect of this process, which is the right information. To take this concept a step further, this dissertation proposes that automated and validated interpretations of data derived from computational models are necessary to obtaining the "right information" and for improving the clinical decision-making process in the 21st century.

The need for an automated interpretation of clinical data is obvious: the availability of medical data continues to grow exponentially while a clinician's available time continues to shrink. Current estimates for the rate of growth of healthcare data is in the range of pentabytes (i.e. 10^{15} bytes) per year, while clinicians surveyed between 2013-2016 spent, on average, 15-25 minutes face-to-face with their patients.(5; 6) Simply put, no clinician has the time to synthesize important findings from millions of patients across thousands of variables in the short time they have available for assessments.

Currently, there exist two schools of thought in terms of how to interpret data through modeling systems: 1) using models for knowledge discovery and hypothesis generation, and 2) using models to test hypotheses in a simulated fashion. The former is the goal of machine learning models, which ultimately attempt to develop a function-approximate view of the world by, at their core, leveraging the shared information between a predictor and the outcome of interest.(7; 8) The latter is the goal of mechanistic models, which are built on a solid prior understanding of how the world, or in this case human body, functions to study how an outcome will change when a system is perturbed.(9) While often viewed as separate, this dissertation proposes novel

methods of both modeling paradigms applied to difficult clinical questions, specifically in examples of pulmonary arterial hypertension.

Background: Pulmonary Arterial Hypertension

Pulmonary arterial hypertension (PAH) is a debilitating disorder of the pulmonary vasculature.(10) It falls under a subcategory of pulmonary hypertensive disorders (Group 1 of five different groups), the rest of which have unique diagnoses and are outside the scope of this dissertation, but are described briefly in Table 1.1.

Epidemiological Group	Description
1	• Primary to pulmonary artery • Includes idiopathic, heritable, drug-induced etiologies, or those associated with connective tissue disorder or congenital heart disease
2	• Caused by left heart disease, resulting in elevations in mean pulmonary arterial pressure or pulmonary capillary wedge pressure or both
3	• Caused by lung disease or hypoxemia
4	• Caused by chronic pulmonary thromboembolism (CTEPH)

Table 1.1: Epidemiological Descriptions for Types of Pulmonary Hypertension.

Specifically, Group 1 PAH is diagnosed as hypertension that is isolated to the pulmonary arteries (precapillary) and requires the following hemodynamic measures (made through right heart catheterization): 1) a mean pulmonary arterial pressure of > 25 mmHg (although, diagnosis can be made at pressures as low as 20 mmHg, but this remains in contention), 2) a pulmonary capillary wedge pressure of < 15 mmHg, and 3) pulmonary vascular resistance of > 3 mmHg/(L/min)).(10) Progressive narrowing of the pulmonary artery due to both cellular proliferation and vasoconstriction leads to increased

afterload on the right ventricle of the heart, causing the right ventricle to hypertrophy and eventually fail. There is no known cure for this condition.(10)

Current conservative estimates of global prevalence of PAH are 15 cases per million, with an incidence of 15-50 million people in the United States and Europe, giving it the unique status of being a “rare” or “orphan” disease. Within the classification of Group 1 PAH, idiopathic and heritable etiologies are the most common, followed by multiple associated forms related to drug use, human immunodeficiency virus (HIV), and connective tissue disorders.(10)

The cellular pathophysiology of PAH is an incomplete but complex and active area of research, involving genetics, epigenetics, and environmental triggers that lead to pulmonary vascular remodeling.(11) Multiple cells of the pulmonary vasculature (e.g. fibroblasts, smooth muscle cells, myofibroblasts, etc.) are implicated in the progression of vascular remodeling, but much of the therapeutic focus has been on pulmonary endothelial cells. Three major cellular pathways specific to pulmonary endothelial cells become dysfunctional throughout the course of the disease: 1) endothelin, a peptide family of vasoconstrictors, 2) nitric-oxide-cyclic GMP (NO-cGMP), a signaling cascade that affects smooth muscle relaxation and vasodilation, and 3) prostacyclin, a signaling molecule that inhibits smooth muscle cell proliferation and fibroblast growth, as well as enhancing vasodilation. The disruption of these pathways results in imbalanced endothelial factor release that causes pulmonary arterial vasoconstriction and cellular proliferation. All three pathways are currently targeted by pharmaceutical treatments.(10; 11)

Treatment Guidelines and Clinical Management of PAH

Currently, most clinical treatment options specific to PAH focus on pharmaceutical vasodilation of the pulmonary artery and/or reduction of cell proliferation to slow vascular luminal thickening. Over 14 pharmaceutical treatments specific to the treatment of PAH have been approved by the U.S. Food and Drug Administration (FDA) since 1995, yet the growing number of treatment options has not conferred a dramatic improvement in prognosis.(12) With modern treatment strategies, current estimates of survival are nearly double what they were 20 years ago but remain devastating – median survival is now 7 years versus 3 years in 1991.(13)

All PAH medications have significant tradeoffs between efficacy and side effects; further, nearly half of these medications are approved based only on their ability to improve exercise capacity in patients.(10; 12) More recent PAH drug efficacy clinical trials have focused on therapeutic improvements in delaying time to clinical worsening, a composite endpoint with varying definitions that typically includes time to first (all-cause) death, hospitalization due to worsening PAH, or disease progression (measured as an increase in right heart failure symptoms and a reduction in exercise capacity).(12) Although the clinical relevance of such a composite endpoint is clear, there remains significant debate around how best to define clinical worsening and its relevance to the patients' desires for medications that not only delay their death, but also improve their quality of life. Indeed, side effects of certain medications can result in a significant reduction in mortality but reduce quality of life drastically.

Understanding each drug class and its risk-benefit tradeoffs is important for providing further context to the challenges of clinical decision science for PAH. As stated,

currently three major pathways are targeted by PAH medications: endothelin, NO-cGMP, and prostacyclin.

Targeting of the endothelin pathway is accomplished with endothelin receptor antagonists (ERAs), which may target one or both receptor types (A and/or B). Ambrisentan, Bosentan, and Macitentan are all oral ERAs approved specifically for the treatment of PAH, and all have demonstrated an improvement in time to clinical worsening and exercise capacity in randomized, placebo-controlled clinical trials.(14-16) Depending on the location of clinical care, treatment with Ambrisentan and Bosentan require or strongly recommend regular liver function testing due to increased risk of liver damage. Macitentan can cause significant reductions in blood hemoglobin, so clinical monitoring of red blood cell count is required.(10)

Targeting of the nitric oxide cyclic-GMP (NO-cGMP) pathway is accomplished with phosphodiesterase type 5 inhibitors (PDE-5i) or soluble guanylate-cyclase (sGC) stimulators.(11) Sildenafil, Tadalafil, and Vardenafil, are all oral PDE-5is approved specifically for the treatment of PAH, with significant pulmonary vasodilation effects.(17-19) Riociguat is the only sGC currently approved for PAH.(20) Clinical trials for these medications have been short (12-16 weeks) and primarily focused on improvements in exercise capacity, with some favorable results shown for time to clinical worsening during these short-term trials, but no statistical significance.(17-19; 21) However, vasodilatory effects of these medications are short-lived and medications must be taken two or three times a day. Side effects, including syncope, flushing and epistaxis (i.e. nosebleeds), are largely mild to moderate, though hypotension is a significant concern at high dosages.

Further, although combination therapy is desirable for synergistic effects, the combination of sGCs and PDE-5i is contraindicated due to the risk of significant hypotension.(10; 22)

Finally, the prostacyclin pathway is targeted through stable analogues of the prostacyclin compound (i.e. prostanoids) as well as through prostacyclin receptor agonists.(11) These medications come in a wide variety of delivery routes, including oral, subcutaneous, intravenous, and inhaled aerosols.

Intravenous prostacyclin analogues include epoprostenol, the first medication indicated for PAH, and Iloprost.(23; 24) Epoprostenol is considered the most effective PAH treatment and is still the only monotherapy (i.e. single treatment) to demonstrate a significant reduction in mortality.(10; 23; 25) However, epoprostenol requires continuous infusion, a permanently tunneled central venous catheter, and continuous cooling. Intravenous Iloprost is more shelf-stable at room temperature and appears to be equally as effective but this has not been demonstrated in a randomized controlled trial.(26-28) All intravenous, continuous infusion prostanoids come with significant risk of infection (including sepsis), catheter obstruction, significant reductions in patient quality of life, and substantially higher financial cost.(29; 30) Further, pump malfunctions that result in abrupt interruption of infusion can result in pulmonary hypertension rebound, with a significant risk of sudden RV failure and mortality.(10) Therefore, treatment with intravenous epoprostenol is typically reserved for the most severe cases of PAH. More recently, PAH centers in Japan have shown significant survival benefits with early i.v. epoprostenol treatment, regardless of disease severity, but this patient population is small.(31)

Subcutaneous delivery of the prostanoid Treprostinil is available, allowing for a substantial reduction in risk of systemic infections, but has not been determined to be as

effective as intravenous treatment. Further, subcutaneous injection still comes with significant local injection site pain (patients describe this as an intolerable burning sensation) and reductions in patient quality of life.(30; 32) Oral delivery of prostanoids (Treprostinil and Beraprost) are effective primarily for improvements in exercise capacity but poorly tolerated, and commonly results in headaches, diarrhea, flushing, and jaw pain.(33; 34) Aerosol administration of prostanoids is also available and better tolerated, but requires 6-9 inhalations per day.(24; 35; 36) The only oral medication of prostacyclin receptor agonist, Selexipag, is a relatively well-tolerated treatment that has demonstrated significant improvements in time to clinical worsening in a randomized controlled trial, but still has not been shown to significantly reduce mortality.(37)

Overall, no pharmaceutical treatment for PAH can fully reverse or halt vascular remodeling or disease progression. Combination therapies are an attractive strategy for targeting multiple disrupted vasoconstrictive and mitogenic pathways, but come with increased risk of drug intolerance/sensitivity and other more severe side effects.(10) Further, studies of upfront combination therapy with epoprostenol have not shown significant benefits in time to clinical worsening.(38; 39) There is evidence to suggest a significant benefit in exercise capacity and quality of life using sequential add-on of oral, intravenous, or inhaled therapies(35; 36; 40; 41), but only one randomized controlled clinical trial has demonstrated a benefit in using combination oral therapy upfront (i.e. for treatment-naïve patients).(42) A preliminary pilot study demonstrated significant improvements in exercise capacity, hemodynamics, and symptoms of right heart failure for severe patients on upfront triple combination therapy, but these results have not yet been demonstrated in a randomized clinical trial.(43)

Other strategies for clinical management of PAH progression and symptoms include the use of diuretics to reduce venous congestion resulting from progressive RV failure and oxygen administration.(10) Calcium channel blockers (CCB) are also prescribed, but only for idiopathic PAH patients deemed “vasoreactive”, meaning the patients have a significant, acute reduction in mean pulmonary arterial pressure during inhaled administration of nitric oxide. This constitutes only an estimated 10% of IPAH patients, roughly 0.5% of Group I PAH patients overall, but these patients do remarkably well on CCB therapy.(44; 45) Other commonly used off-label treatment strategies, such as oral anticoagulant therapy, do not have clear evidence of improving clinical outcomes.(10) Finally, there is still a significant degree of interpatient variation in disease progression, despite modern-day clinical management strategies. Seemingly stable PAH patients can rapidly decompensate for reasons unknown to clinicians.(46)

Causes of rapid decompensation in PAH patients are not well understood but have been linked to ischemia in the left and/or right heart, especially caused by compression of arteries by enlarged pulmonary arteries, as well as ventricular and supraventricular arrhythmias.(47) Despite this, a general strategy of treating with antiarrhythmics, has not shown improvements in PAH clinical outcomes, though targeted therapies for patients with arrhythmias as a comorbidity is recommended. Further, an estimated 50-60% of PAH patients die from other systemic causes that are related to or worsened by their pulmonary hypertension, including acute renal failure (can be brought on by right ventricular dysfunction, which causes renal hypoperfusion and venous backflow), sepsis (in some cases, caused by the very intravenous treatment intended to manage the patient’s PAH), or respiratory diseases such as interstitial lung disease or pneumonia.(48) Therefore, to

improve clinical outcomes, there is a critical need to identify patients who are most vulnerable to acute stressors, such as infection and physical exertion, allowing for timely intervention and treatment elevation.

Current Methods for PAH Prognosis and Risk Estimation

As knowledge of the mechanisms that lead to early death are currently limited, optimization of a PAH treatment plan cannot yet be assessed on a causal basis. Due to this knowledge gap, treatment guidance through risk assessment is considered the best clinical strategy at this time.(10) The 2015 European Respiratory Society guidelines provide loose recommendations for risk assessment, based on a number of clinical variables, which is reproduced in Table 1.2.

Determinants of prognosis (estimated 1-year mortality)	Low Risk (<5%)	Intermediate Risk (5-10%)	High Risk (<10%)
Clinical signs of right heart failure	Absent	Absent	Present
Progression of symptoms	None	Slow	Rapid
WHO Functional Class	I,II	III	IV
Six Minute Walk Distance (m)	> 440	165-440	< 165
Cardiopulmonary exercise testing			
Peak VO_2 (mL/min/kg) (or % predicted)	> 15 (>65%)	11-15 (35-65%)	< 11 (<35%)
VE/VCO_2 slope	< 36	36-44.9	>45
Circulating Biomarkers			
Plasma Brain Natriuretic Peptide (ng/L)	<50	50-300	>300
Plasma NT-proB-type Natriuretic Peptide (ng/L)	<300	300-1400	> 1400
Imaging (Echocardiography)			
Right atrial area (cm^2)	< 18	18-26	26
Presence of pericardial effusion	None	None or Minimal	Present
Hemodynamics			
Right atrial pressure (mmHg)	< 8	8-14	> 14
Cardiac Index ($L/min/m^2$)	> 2.5	2.0 – 2.4	< 2.0
Mixed venous oxygen saturation (%)	> 65	60-65	<60

Table 1.2: 2015 European Respiratory Society Guidelines for Pulmonary Arterial Hypertension Risk Assessment.

Briefly, ERS risk assessment depends on echocardiographic, hemodynamic, and circulating biomarkers, as well as functional metrics of the patient's symptoms and exercise capacity. Within the echocardiographic parameters, right atrial area is a predictor of mortality, as increases in area demonstrate increased venous congestion and worsening right heart failure.(49). Pericardial effusion, or fluid within the pericardium, is associated with increased mortality in PAH, but the causal link is not well understood.(50) Within hemodynamic parameters, cardiac index, which is cardiac output normalized by body surface area, indicates right heart function in terms of perfusion, while right atrial pressure indicates right heart function in terms of venous congestion. Mixed venous oxygen saturation indicates when cardiac index is insufficient, as it demonstrates a reduced capacity to meet tissue oxygenation requirements and may be more informative than cardiac index itself.(51) With regards to circulating biomarkers, brain natriuretic peptide (BNP) is a cardiac neurohormone that indicates heart stress and increased pressure, resulting in a series of physiological effects that attempt to reduce heart strain via the renal, adrenal, vascular and cardiac systems. BNP's prohormone form, N-terminal pro brain natriuretic peptide (NTproBNP), while inactive, can also be used to indicate strain on the heart.(52)

Finally, both maximal and submaximal tests of exercise capacity, including the six-minute walk test and cardiopulmonary exercise test (CPET), are used as predictors of mortality and further indicate heart function. The six-minute walk test is extremely convenient to administer, requiring only for a patient to walk a 3-meter track repeatedly for the six-minute time period. However, it is somewhat contentious in its clinical relevance due to its submaximal nature – a patient's effort level can vary, and some

patients require oxygen to complete the test at all, leaving interpretation of the distance unclear. CPET, a maximal exercise test, is less convenient to administer, but provides two predictive values of mortality: peak oxygen uptake (VO_2), which is lower in higher risk PAH patients due to lower cardiac output, and minute ventilation to carbon dioxide production (VE/VCO_2), which is higher in higher risk patients, due to alveolar hypoperfusion.(53)

Generally, no single measure from these recommended variables is considered perfect and the guidelines recommend measuring multiple clinical variables to assess risk but give no indication of how many or which variables are the most crucial. Further, there is a severe lack of consistency for which clinical variables are monitored, given their relative expense and convenience. Cardiopulmonary exercise testing and echocardiography are rarely used for routine assessment due to expense and inconvenience, and hemodynamics are measured on long time scales due to their invasive nature.

These recommendations are meant to guide clinicians in identifying if their patient is a low, intermediate, or high risk of mortality within the next year. Then, given an identified risk level, clinicians are meant to follow the treatment guidance flowchart, shown in Figure 1.1.(10) There are no specific recommendations for how to treat PAH patients with one or more comorbidities, which is common, and there is evidence to suggest that risk assessment tools are less reliable in these patients.(54)

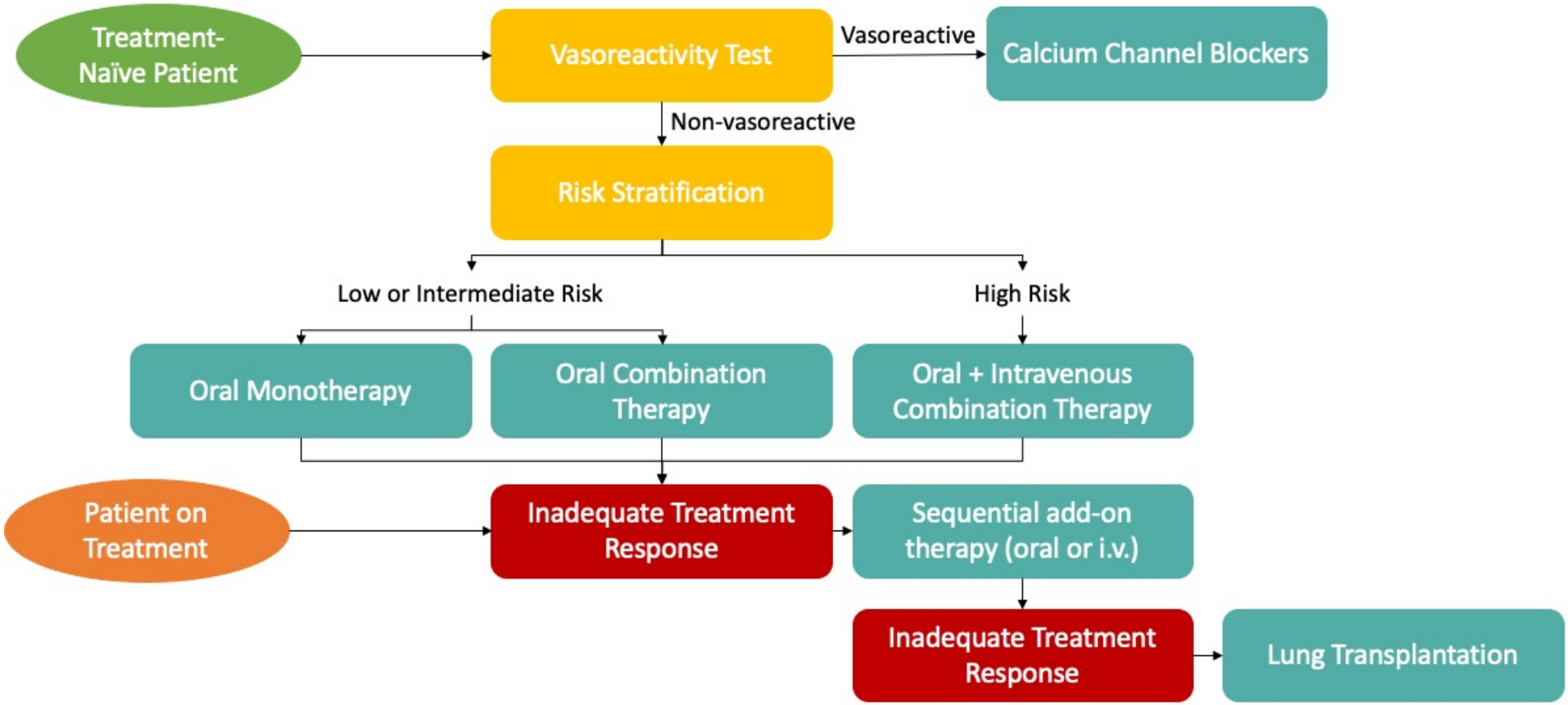


Figure 1.1: 2015 European Respiratory Society Treatment Algorithm.

While the guidelines maintain that a comprehensive assessment of multiple clinical measures is required for accurate risk estimation, the recommendations overall do not recommend a more specific or quantitative strategy and they specify that risk estimation is still largely up to the clinician's expert opinion. In fact, many clinicians eschew multivariate risk assessment altogether and rely on their own individual clinical experiences or entirely on a qualitative assessment of the patient's heart failure symptoms, known as their New York Heart Association (NYHA) Class (Table 1.3).(55)

New York Heart Association Class	Patient's signs and symptoms
I	No limitation of physical activity. Ordinary activity does not cause undue fatigue, palpitation, or dyspnea (shortness of breath).
II	Slight limitation of physical activity. Comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnea (shortness of breath).
III	Marked limitation of physical activity. Comfortable at rest. Less than ordinary physical activity causes fatigue, palpitation, or dyspnea.
IV	Unable to carry on any physical activity without discomfort. Symptoms of heart failure at rest. if any physical activity undertaken, discomfort increases.

Table 1.3: New York Heart Association Functional Classification.

NYHA class is defined on a very simple scale of I to IV, where NYHA class I patients are considered the least severe (with greatest functional and exercise capacity) and class IV are considered the most severe (with extremely poor functional and exercise capacity). However, it has been demonstrated in several studies that NYHA class alone is too broad to determine treatment strategies, and as a result, patients in class II and III are likely to be both undertreated and overtreated.(56; 57)

Further, the treatment guidance flowchart, as reproduced in Figure 1, is nonspecific about upfront treatment strategies for low and intermediate risk patients (suggesting either oral combination therapy or monotherapy, as the clinician sees fit). In 2018, *Frost et al* determined through a retrospective analysis that patients of all risk levels equally benefit from upfront oral combination therapy, but used a risk tool (REVEAL [1.0], described in detail below) that is now considered outdated.(58) Ideally, risk assessment would be conducted in a quantitative, serial fashion to determine if treatment response is adequate, but the guidelines are also nonspecific on what constitutes an adequate clinical response to treatment.

In lieu of a standardized quantitative risk assessment, several international researchers have developed risk tools based on the ERS guidelines to create "risk calculators". These calculators aim to stratify patients into a more accurate low-, intermediate-, or high-risk group based on an estimated probability of mortality within the next year. Current popular tools include three risk calculators developed based on registry data: Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension (COMPERA)(59), the French pulmonary hypertension registry score (French)(60), and the United States Registry to Evaluate Early and Long-Term PAH Disease Management (REVEAL [1.0] and REVEAL 2.0)(61; 62). Each calculator is described in detail below. A comparison of each calculator in terms of variables used, clinical values to determine effect on predicted risk (improves/lowers risk, worsens/increases risk, or greatly worsens/greatly increases risk) is shown in Table 1.4. Generally, the cut-offs for these values are very similar across risk calculators and have

been derived subjectively, not optimized with a more objective or quantitative methodology.

Clinical Variable	Effect on Predicted Risk	COMPERA	French	REVEAL	REVEAL 2.0
mRAP (mmHg)	Improves	<8	< 8	-	-
	Worsens	8-14	-	≥20	≥20
	Greatly Worsens	>14	-	-	-
WHO Functional Class	Improves	I/II	I/II	I*	I
	Worsens	III	-	III	III
	Greatly Worsens	IV	-	IV	IV
Six Minute Walk Distance (m)	Improves	>440	>440	≥440	≥320*
	Worsens	165-440	-	< 165	< 165
	Greatly Worsens	<165	-	-	-
NTproBNP (ng/L)	Improves	<300	<300	-	<300*
	Worsens	300-1400	-	-	-
	Greatly Worsens	>1400	-	-	≥ 1100
BNP (ng/L)	Improves	<50	<50	< 50	< 50*
	Worsens	50-300	-	> 180	≥ 200-800
	Greatly Worsens	>300	-	-	≥800
Cardiac Index ($L/min/m^2$)	Improves	≥ 2.5	≥ 2.5	-	-
	Worsens	2.0-2.4	-	-	-
	Greatly Worsens	<2.0	-	-	-
Mixed Venous Oxygen Saturation (%)	Improves	>65	>65	-	-
	Worsens	60-65	-	-	-
	Greatly Worsens	<60	-	-	-
Pericardial Effusion	Worsens	-		Present	Present
PVR (Wood units)	Improves	-	-	-	< 5
	Greatly Worsens	-	-	> 32	-
(Sitting) Heart Rate (bpm)	Worsens	-	-	>92	>96
Systolic Blood Pressure (mmHg)	Worsens	-	-	<110 mmHg	<110 mmHg
eGFR ($mL/min/1.73\ m^2$ or renal insufficiency)	Worsens	-	-	Insufficiency present	< 60 or Insufficiency present
% predicted DLCO	Improves	-	-	≥ 80	-
	Worsens	-	-	≤ 32	< 40
All-cause hospitalization	Worsens	-	-	-	In last 6 months
Age + Gender	Greatly Worsens	-	-	Male and > 60 yrs	Male and > 60 yrs
PAH Etiology	Worsens			Associated with Connective Tissue Disorder	
	Greatly Worsens			Portal or Heritable Pulmonary Hypertension	Portal† or Heritable Pulmonary Hypertension

Table 1.4: Comparison of Contemporary Pulmonary Arterial Hypertension Risk Scores.

COMPERA was developed in 2017, using data from the Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension registry.(59) The score uses six variables: cardiac index, New York Heart Association functional class, mean right atrial pressure, N-terminus pro brain natriuretic peptide (NTproBNP) or its active hormone product, brain natriuretic peptide (BNP), mixed venous oxygen saturation, and six-minute walk distance. For each variable, patients are given a score of one (1), two (2), or three (3), where three indicates the highest risk for that variable and one indicates the lowest risk for that variable. The average, rounded score (round-up at >= 0.5, round-down at < 0.5) across all variables is then taken for the overall score. Therefore, patients can have an overall COMPERA risk score of one (low risk), two (intermediate risk), or three (high risk). In its original publication, the score was validated using Kaplan-Meier curves and a log-rank test to determine statistically significant differences between low, intermediate, and high risk. Further, COMPERA has been validated in the United States Registry to Evaluate Early and Long-Term PAH Disease Management observational study (i.e. REVEAL registry), with a reported c-statistic of 0.62 (N = 2529).(61)

The French score was developed in 2017, using data from the French pulmonary hypertension registry.(60) The score uses four variables: cardiac index, mean right atrial pressure, New York Heart Association Class, and six-minute walk distance. Patients are ranked on a scale from zero (0) to four (4), with one-point increments, where the highest score indicates the least risk (i.e. greatest number of low-risk features) and lowest score indicates the most risk (i.e. fewest number of low-risk features). The original publication of the score does not specifically recommend a simplified grouping for low, intermediate, and high risk, but suggests that survival between those with a score of three and score of

four may not have significant differences in survival. Original validation of the score demonstrated that it was a significant predictor of transplant-free survival in a Cox proportional hazards model both at time of PAH diagnosis (N = 1591, $p < 0.001$) and at first evaluation within one year (N = 1017, $p < 0.001$). The model has been notably validated in the following retrospective studies: 1) REVEAL registry, as a predictor of one-year mortality (c-statistic = 0.64, N = 2529),(61) 3) the Assessing the Spectrum of Pulmonary Hypertension Identified at a Referral Center MRI database (ASPIRE) as a predictor of one-year mortality (c-statistic = 0.697, N = 219).(63) The model has been extended to use an optional fifth variable, mixed venous oxygen saturation, and has been converted to a non-invasive form that uses no hemodynamics, but instead uses the following three variables: NTproBNP, NYHA functional class, and six minute walk distance. However, for the sake of comparing scores within this dissertation, only the original French score, which has been validated more thoroughly, will be used.

REVEAL [1.0] was developed in 2015, using data from the United States Registry to Evaluate Early and Long-Term PAH Disease Management observational study.(62) A depiction of the calculator is shown in Figure 1.2.

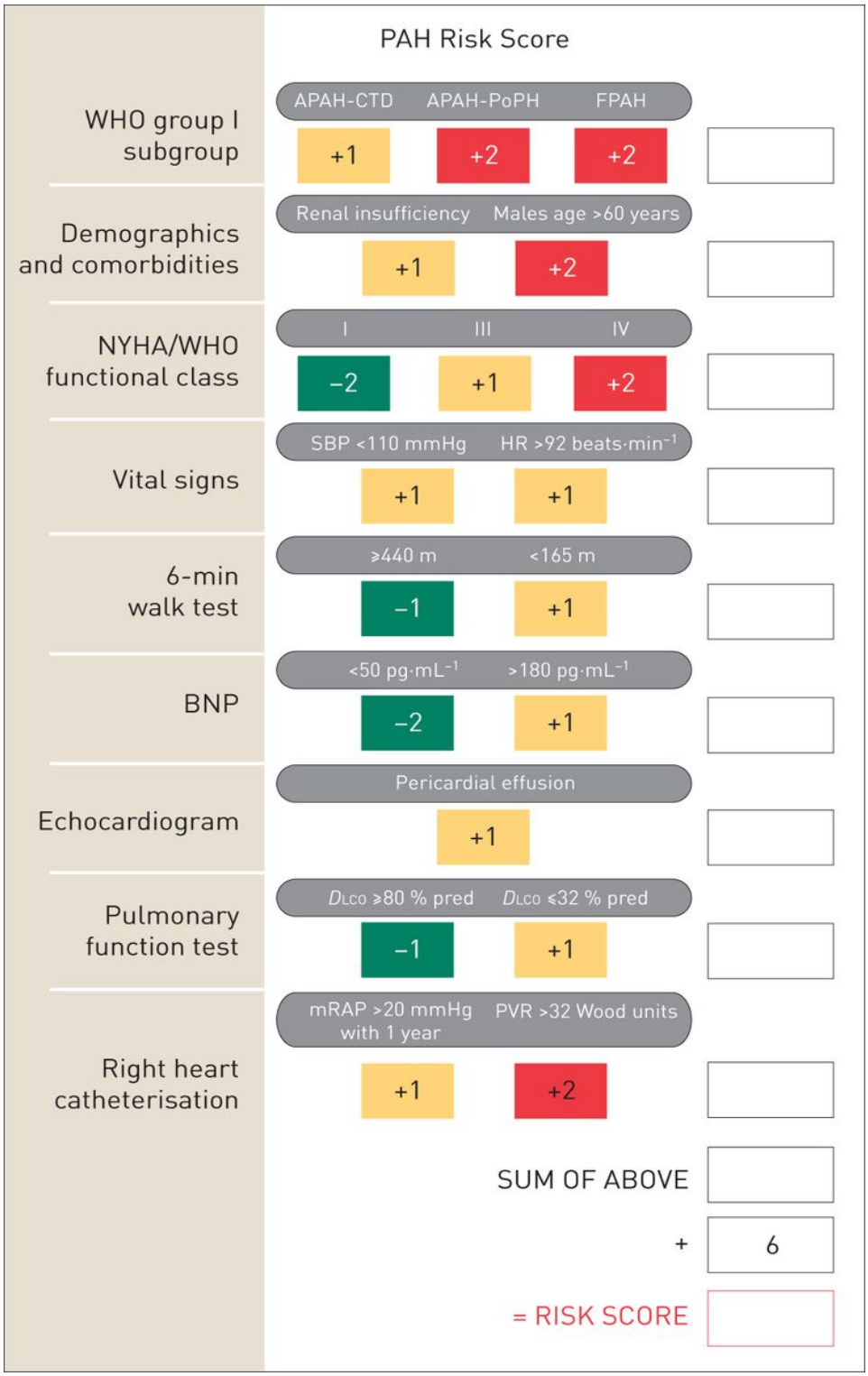


Figure 1.2: REVEAL [1.0] Risk Calculator for Pulmonary Arterial Hypertension. Courtesy of *Raina and Humbert*[66]*(62)*

It was developed by fitting a multivariate Cox proportional hazards regression to one-year survival data and then converting the exponential of the hazard ratio into a linear, additive equation (see Equation 1 in section *Risk Modeling Methodology* for further details). Cut-points for each variable are identified to either add or subtract a specific number of points from the overall risk score, based on the exponential value of the hazard ratio. REVEAL [1.0] uses 13 clinical variables: sex, age, clinically identified "renal insufficiency", New York Heart Association functional class, systolic blood pressure (mmHg), sitting heart rate, six-minute walk distance, NTproBNP or BNP, presence of pericardial effusion, percent predicted diffusing lung capacity for carbon monoxide (D_{LCO}), mean right atrial pressure and pulmonary vascular resistance. Using REVEAL [1.0], patients can receive a risk score of zero (0) to eighteen (18), with one-point increments, where a score of 0 indicates the lowest risk of death within the next year, and a score of 18 indicates the highest risk of death within the next year. According to the original publication, it's appropriate to use REVEAL [1.0] to more simply group patients as low risk (REVEAL [1.0] <= 6), intermediate risk (REVEAL [1.0] 7-8) or high risk (REVEAL [1.0] > 8).

REVEAL 2.0 was developed in 2019, also using data from the United States Registry to Evaluate Early and Long-Term PAH Disease Management observational study.(61) It is an update on the previously published REVEAL [1.0] calculator. A graphical depiction of the REVEAL 2.0 calculator is shown in Figure 1.3.

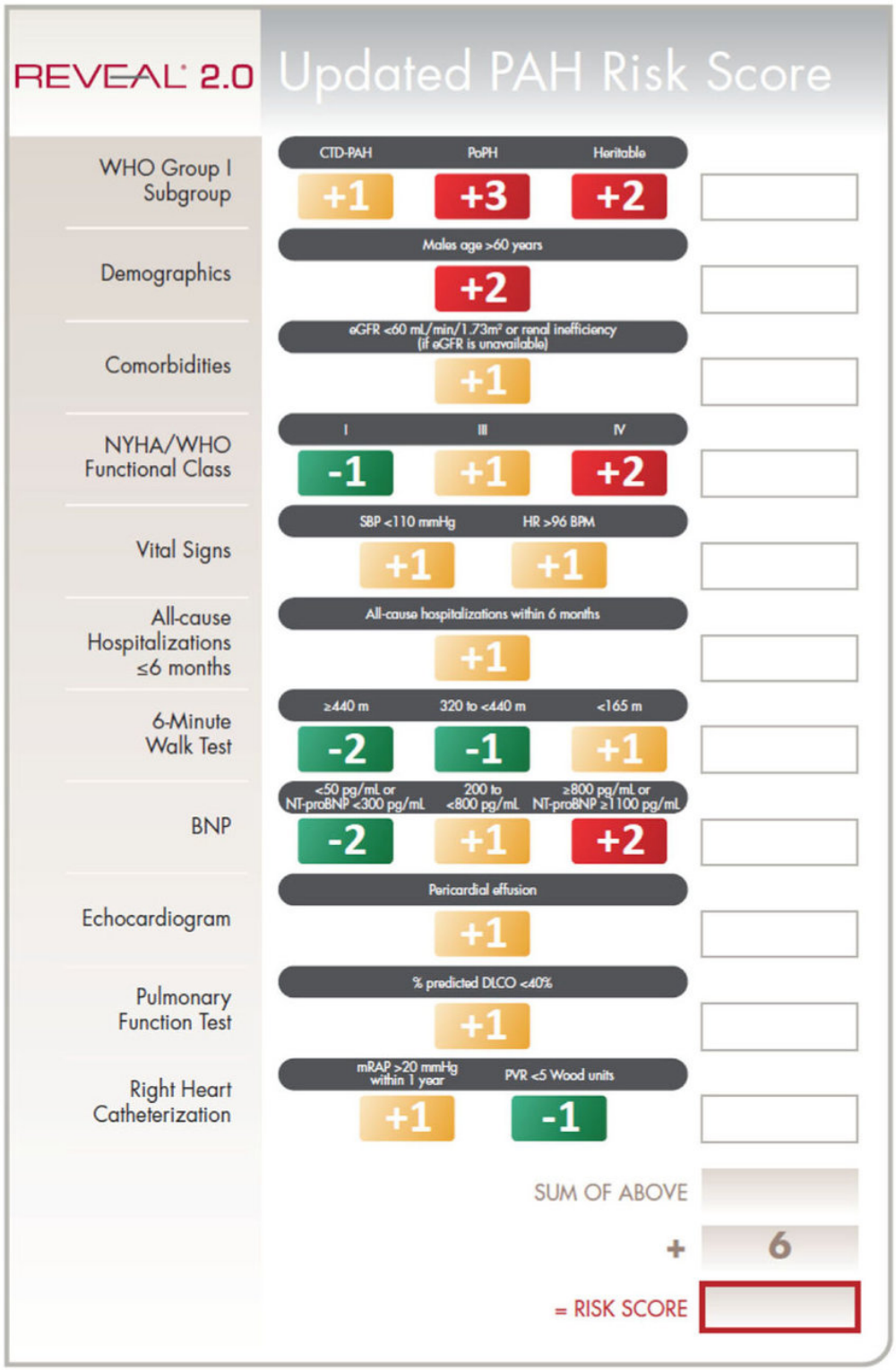
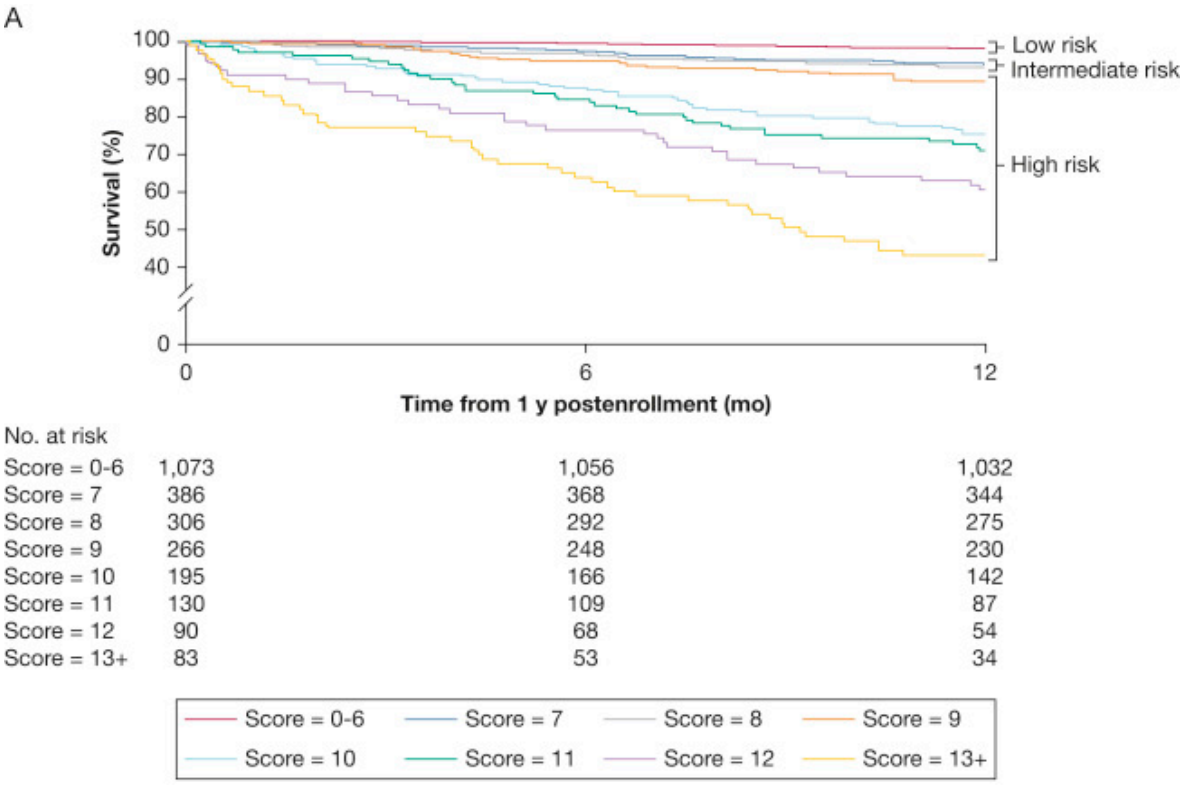


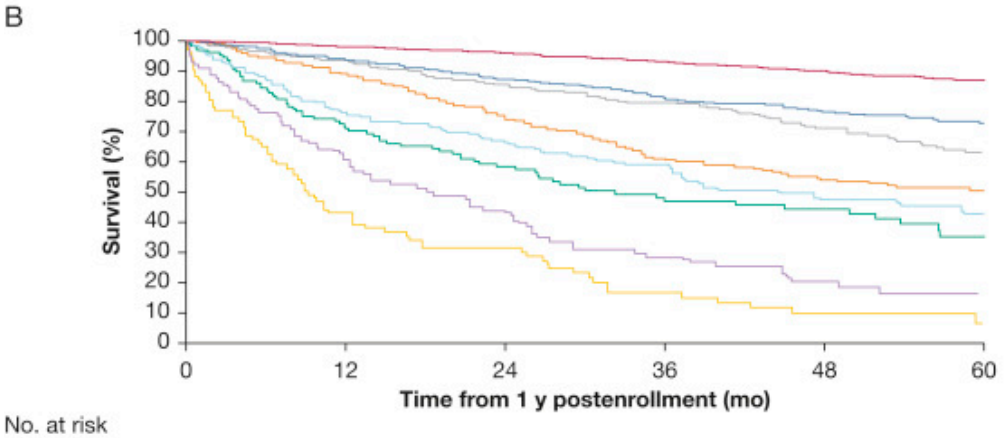
Figure 1.3: REVEAL 2.0 Risk Calculator for Pulmonary Arterial Hypertension. Courtesy of *Benza et al* [61]

Specifically, REVEAL 2.0 uses the same 13 clinical variables as REVEAL [1.0], but with slight differences in the cut-points used and the addition of a fourteenth variable: recent hospitalizations from any cause in the last 6 months. Using REVEAL 2.0, patients can receive a risk score of zero (0) to twenty-three (23), with one-point increments, where a score of 0 indicates the lowest risk of death within the next year, and a score of 23 indicates the highest risk of death within the next year. REVEAL 2.0 is one of the most accurate PAH risk calculators currently used, with a strong ability to discriminate survival differences: patients with a REVEAL 2.0 score of zero have a one-year survival estimate of greater than 95%, while those with scores of more than 13 have a less than 50% chance of survival. One-year and five-year survival curves for each risk value (from a REVEAL 2.0 score of zero to a score of >13) are shown in Figure 1.4.



No. at risk

Score = 0-6	1,073	1,056	1,032
Score = 7	386	368	344
Score = 8	306	292	275
Score = 9	266	248	230
Score = 10	195	166	142
Score = 11	130	109	87
Score = 12	90	68	54
Score = 13+	83	53	34



No. at risk

Score = 0-6	1,073	1,032	961	839	743	343
Score = 7	386	344	304	257	217	114
Score = 8	306	275	236	194	155	79
Score = 9	266	230	178	133	100	46
Score = 10	195	142	116	88	58	32
Score = 11	130	87	67	40	30	13
Score = 12	90	54	34	21	11	5
Score = 13+	83	34	24	10	5	2

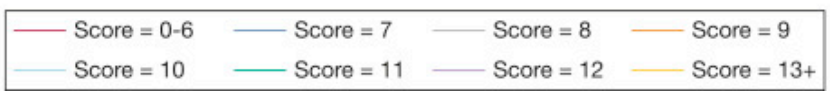


Figure 1.4: Survival Curves (One-year and Five-Year) by REVEAL 2.0 Risk Score. Courtesy of *Benza et al* [61]

According to the original publication, it's appropriate to use REVEAL 2.0 to more simply group patients as low risk (REVEAL 2.0 <= 6), intermediate risk (REVEAL 2.0 7-8) or high risk (REVEAL 2.0 > 8). REVEAL 2.0 has been validated in the REVEAL registry, with a c-statistic of 0.76 (N = 2529) and in the Pulmonary Hypertension Society of Australia and New Zealand (PHSANZ) registry, with a c-statistic of 0.74 (N = 1011).(64)

The Pulmonary Hypertension Outcomes Risk Assessment (PHORA) model was published in 2019 and was the first attempt to design a machine-learning model for PAH risk stratification. The variables and cut-points were taken directly from REVEAL 2.0, and data from the REVEAL registry was used to learn a tree-augmented naïve Bayesian network, which is described in more detail later in the chapter. Briefly, it is a machine-learning model that learns an acyclic directed graph, modeling the impact of both discrete variables in the prediction of the outcome but also the interaction of variables. The network is shown in Figure 1.5.

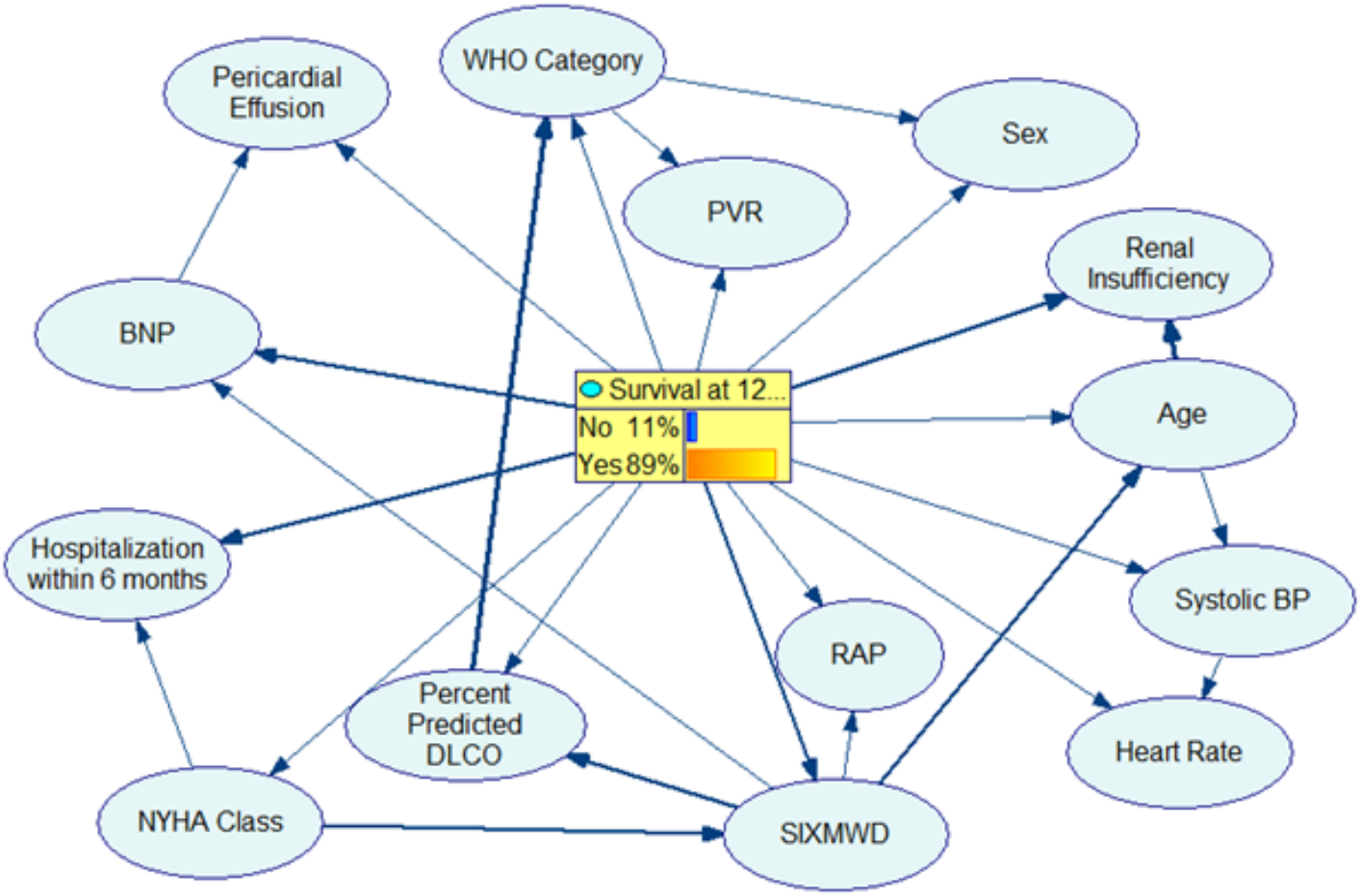


Figure 1.5: Pulmonary Hypertension Outcomes Risk Assessment (PHORA) Tree-Augmented Naive Bayesian Network.

PHORA's cross-validated performance in the REVEAL registry was reported as a c-statistic of 0.80 (versus REVEAL 2.0's c-statistic of 0.76), and it has been validated in the COMPERA registry (c-statistic = 0.74) and Pulmonary Hypertension Society of Australia and New Zealand (PHSANZ) registry (c-statistic = 0.80).(65) To date, there has been no further exploration of optimizing the variables used in the model (especially, reducing required number of variables), optimizing cut-points, or optimizing network structure. Further, there's been no exploration on using the PHORA model for hypothesis generation for novel insight regarding physiological and/or demographic interactions.

Overall, each calculator uses a combination of hemodynamics, demographics, and/or laboratory values to assign a risk level to a patient based on predicted mortality, and is generally calibrated such that low risk reflects less than 5% one-year mortality rate, intermediate risk reflects 5-10% one-year mortality rate, and high risk reflects greater than 10% one-year mortality rate.(59-61) However, there is currently no single, agreed-upon tool to determine risk and all common clinical risk tools based on registry data have limitations in their accuracy (as determined through c-statistic and receiver operator curve analysis) and levels of validation.

Adoption of risk tools, instead of clinical gestalt or NYHA class, is slow for a number of reasons: ambiguity in which tool to use among multiple published tools, low confidence in the (lack of) sophistication of the tool, and an inability to measure all required variables in a timely manner.(66) While a risk tool that considers a large number of variables may be more accurate than most other calculators, time is of the essence for clinicians and tools must be convenient and intuitive to use, yet not overly simplistic. Risk scores that require additional patient testing beyond the standard routine measures are time-consuming, costly, and not always available. Therefore, future risk tools should attempt to reduce barriers to entry by relying on as few variables as possible, more intelligently handling missing variables and improving upon sophistication while avoiding purely "black-box" solutions. Further, to ensure clinical adoption, risk tools should reflect an accurate physiological understanding of PAH mortality and allow for explainable interpretations as to why a patient is scored in one way or another.

Novel Applications of PAH Risk Scoring

In addition to improving treatment strategies, a more accurate risk assessment could be applied to improve many other aspects of PAH clinical science.

First, PAH clinical trials designed to demonstrate drug efficacy have become increasingly burdensome, as so-called “event-driven” trials are required to demonstrate that a drug significantly reduces the progression of the disease.(12) These event-driven trials observe patients over a long period of time (averaging three years or more), waiting for a patient to display worsening symptoms of their PAH, and are statistically powered by the number of events observed overall. To achieve enough events to adequately power a study, upwards of 500 patients must be enrolled. However, there is evidence that suggests enrolling patients based on their risk of clinical worsening would allow more events to be observed in a shorter time frame, improving trial efficiency, and reducing ethical concerns of long-term treatment with placebo.(67) This strategy is referred to as by the U.S. Food and Drug Administration (FDA) as “prognostic enrichment”. Demonstrating the feasibility of this will be the focus of Chapter 2.

Second, as stated in the section, *Treatment Guidelines and Clinical Management of PAH*, it is not well-understood whether PAH patients who have never received a PAH specific therapy (i.e. treatment-naïve) should begin on a regimen of multiple oral drugs that target different cellular proliferation/vasoconstriction pathways (i.e. combination therapy) or would still be well-served on a single oral medication. This question could be answered by a more accurate risk stratified method, as there is a potential benefit-risk tradeoff in using multiple medications to slow the progression of PAH versus

overtreatment and increased exposure to drug side effects. Demonstrating this possibility will be the focus of Chapter 3.

As there are many aspects of PAH clinical decision science that would benefit from an improved prognostic score, this dissertation will now discuss in detail novel methods for improving on PAH risk modeling methodology, comparing and contrasting machine learning techniques versus more traditional statistical risk models that are currently used. Improvement of risk stratification tools with machine learning techniques will be the focus of Chapter 4.

Finally, novel techniques for physiological modeling can also play a role in improving understanding the causal mechanisms by which PAH patients experience disease worsening and higher risk of mortality. This will be the focus of Chapter 5.

Risk Modeling Methodology

As discussed above, there is a clear need for improved prognostic tools for patients with pulmonary arterial hypertension. In this section, current statistical methods (i.e. Cox proportional hazards regression models) will be reviewed and compared to well-established, but newly applied, machine learning methods. Weaknesses and strengths of both are discussed. Specific focus is given to the Bayesian network model using discrete variables, as well as the various methods by which variables are selected to optimize such a model. All of the discussed methods for Bayesian network modeling and feature selection are applied in Chapter 4.

Cox Proportional Hazards Regression Models

Death is an event with obvious time-dependent importance. Predicting if a patient will die in a static, pre-specified timeframe (e.g. six months, one-year, five-years, etc) can be considered a binary event. However, further context can be included in a dynamic time frame by representing the event as a data "tuple" of presence or absence of event and time under observation. Survival analysis is the collection of statistical methods that estimate time-to-event and determine the confidence of the estimation or significant differences between two or more groups in their estimated probability of survival.(68)

Cox proportional hazards regression is a semiparametric method of survival analysis that determines whether a hazard (β) is a significant predictor of the time-to-event, and its effect size is estimated as a coefficient in the following equation:

$$\lambda(t|X_i) = \lambda_0(t)\, exp\left(\beta_1 X_{i_1} + \cdots + \beta_N X_{i_N}\right), \quad (1)$$

for subject *i* at time *t* , λ_0 is the baseline hazard rate, covariate vector (i.e. vector of clinical variables) *Xi* and covariates 1 through *N*. (69)

Cox proportional hazards models can therefore be univariate (modeling a single clinical variable) or multivariate (modeling multiple clinical variables). For the purposes of risk modeling, a variable is said to be an "independent predictor" of the event (i.e. death) if it maintains its statistical significance in a multivariate model among other predictors. However, this is something of a misnomer, as some variables are only significant predictors when present in a model that considers another specific predictor or group of predictors. Therefore, there is no guarantee that a predictor that is significant in a multivariate model is also significant in a univariate model, or vice versa. This is the first

of many limitations in Cox proportional hazards regression, as it is not immediately clear what the true relationships between predictor variables may be.(68)

That said, there are significant benefits in using a parameterized model with such strict assumptions when sample sizes are small. Determining the statistical significance of an interaction requires an estimated 4-10x the sample size required to detect statistical significance of main independent effects, depending on the effect size of the interaction.(70; 71) Cox proportional hazards models are therefore less prone to overfitting in small sample size situations, provide clear intuitive interpretations of the variables that contribute to risk of event, and can make use of censored data.

Bayesian Network Modeling

Bayesian networks are probabilistic, graphical models based on Bayes' theorem, which states that the conditional probability of observing event A given event B (i.e. posterior probability) is equal to the joint probability of observing event A and B, divided by the probability of observing event B(72):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \times P(A)}{P(B)} \quad (2)$$

A Bayesian network therefore uses observed data to estimate the joint probability of events A and B and the marginal probability of observing event B to then later predict the probability of observing event A given the known status of event B. Clinically, this allows a Bayesian network to predict a binary event (i.e. event A), such as death in the next year, based on an observed value of a clinical variable (i.e. event B). Of course, event B itself can represent the joint probability of observing multiple events, as in,

accounting for multiple clinical variables. This is considered a multivariate Bayesian network versus a univariate Bayesian network.

A key design element that must be chosen when learning a Bayesian network is accurately and efficiently modeling the joint probability of survival and the observed clinical variables. If the clinical variables are all assumed to be conditionally independent given the survival event, the simplest Bayesian network can be developed: the naïve Bayesian network. Graphically, a naïve Bayesian network is depicted as shown in Figure 1.6.

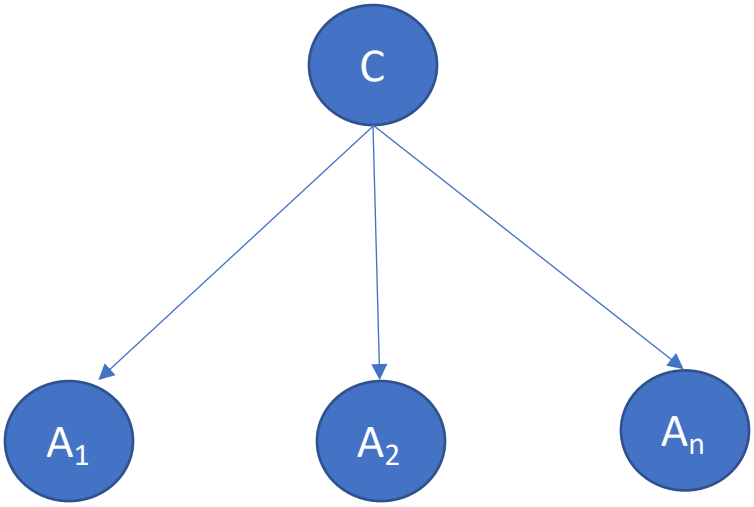


Figure 1.6: Depiction of Naive Bayesian Network.

In Figure 1.6, the node labeled "C" indicates the parent node of the network. Within the context of a clinical prediction model, node C would indicate the predicted outcome of interest, such as mortality within the next year. Nodes labeled "A_1 "through "A_n" indicate variables used to predict the outcome, which are considered conditionally independent of one another based on the parent node C. In other words, a naïve Bayesian network assumes that the observation of any variable is based entirely on the outcome, and

variables do not affect one another otherwise. This can also be described mathematically by applying Bayes' theorem, as shown in Equation 3:

$$P(C|A_1 \dots A_n) = \frac{P(C)\prod_{i=1}^{n} P(A_i|C)}{P(A_i \dots A_n)}, \quad (3)$$

where *C* is again, the outcome of interest, A_i is the i^{th} observed clinical variable, $P(C|A_1 \dots A_n)$ is the conditional (posterior) probability of the outcome given observed clinical variables, *P(Ai|C)* are the conditionally independent probabilities of *Ai* given the outcome *C*, *P(Ai..An)* is the marginal probability of all observed clinical variables, and *P(C)* is the prior probability of the outcome.

Given that clinical variables reflect physiological phenomena and the existence of several interdependencies between organ systems, a naïve Bayesian network is likely a poor choice for risk modeling assuming an adequate sample size. In fact, a clinically relevant example of the weakness of a naïve Bayesian network is provided in the original publication on augmented naïve Bayesian structures, *Friedman et al*, in which a classifier for the diagnosis of diabetes is discussed.(72) Assume that the probability of a patient having diabetes is high if their glucose levels are high and their insulin levels are low. However, in rare instances, a patient may have diabetes but have a normal glucose level and a normal insulin level. In this situation, the naïve Bayesian network will greatly underestimate the probability of a diabetes diagnosis, as it assumes that two independent and unlikely observations given a diabetes diagnosis have been made. In actuality, insulin levels and glucose levels are linked. It is not that two independent unlikely observations have been made: it is more akin to saying that one unlikely observation given a diabetes diagnosis has been made. Allowing for connectivity between these two variables, glucose and insulin, will avoid this underestimation. Therefore, the use of more sophisticated

versions of Bayesian networks provides a significant opportunity to improve clinical modeling of physiological systems.

Augmented naïve Bayesian networks can learn relationships between clinical variables based on observed data, rather than manual implementations of interactions (such as those required in Cox proportional hazards models). The differences then lie in the tuning of hyperparameters (such as the number of interactions one variable may have with other variables) and the algorithm used for learning the interactions. Though there are many more versions of Bayesian networks, the focus here now will be on two types that are classically used for the task of prediction: tree-augmented Naive Bayesian networks, and unrestricted augmented Naive Bayesian networks. Further, two additional Bayesian network structure learning methods used for the task of feature selection will be briefly discussed: Greedy Thick-Thinning Bayesian Networks and PC Bayesian Networks.

Augmented Naive Bayesian Networks

Augmented naïve Bayesian networks (ANBNs) are automated extensions of the naïve Bayesian network model that aim to learn interactions between variables.(72) There are multiple versions of ANBNs, including tree-augmented and unrestricted augmented Bayesian networks, each with different means of learning variable interactions and specific restrictions for the number of interactions each variable can have. Learning variable interactions can increase both the accuracy and knowledge discovery value of a naïve Bayesian network, but they come at a cost of potentially overfitting training data and consequently, performing poorly on unseen data. Therefore, efforts should be taken

to minimize the number of interactions needed to improve model accuracy by choosing the “sparsest” model (i.e. fewest connections) that performs best in cross-validation.

Tree-augmented naïve (TAN) Bayesian networks are a specific extension of naïve Bayesian networks that automate the process of learning interdependencies through the Chow-Liu algorithm, which initially finds the maximum spanning tree to link together all independent variables (i.e. features) prior to learning their relationship with the dependent variable (i.e. label or clinical outcome). A maximum spanning tree is a graphical structure (depicted in

Figure 1.7, in red) composed of a root, edges, and nodes, in which the edges are chosen to maximize their sum and create an undirected structure where each node has exactly one parent.

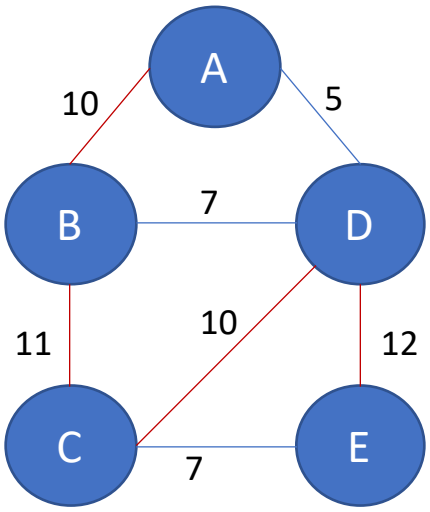


Figure 1.7: Example of a Maximum Spanning Tree.

Once the maximum spanning tree is learned, one variable is chosen at random to be the root node, by which all directed connections now point away, to ensure an acyclic graph. The conditional probability relationships between the predicted outcome and the

variables are now learned, by which variables that are not connected via the maximum spanning tree are independent conditioning on the outcome.

TAN Bayesian networks are somewhat inflexible in learning interactions, as they require that each variable have at least two parents (except the root node), regardless of the relative strength of the interactions. Therefore, to improve model generalizability, Bayesian search can instead be used to augment a Naïve Bayesian network. This method introduces connections between variables based on improvement of the minimum description length (MDL), a score which links variables with high mutual information but also penalizes high connectivity (Equation 4).

$$MDL(B|D) = \frac{log\, N}{2}\,|B| - \sum_{i=1}^{N} log\big(P_B(\boldsymbol{u}_i)\big), \quad (4)$$

where B is a Bayesian network, $|B|$ is the number of parameters of the Bayesian network, D is the training set of N observations of covariate vectors $\boldsymbol{u} = <\boldsymbol{u}_1 \ldots \boldsymbol{u}_N>$, P_B is the probability of the network B given the dataset. Therefore, Equation 4 is minimized when the likelihood of the network is maximized while the number of parameters (i.e. number of connections between nodes) needed to describe the network is minimized. This method doesn't enforce the two-parent rule, as instead the maximum number of parents is set as a hyperparameter by the model developer. This increased flexibility in connectivity can allow for both sparser models (at most two parents, but not requiring two parents) or much more connected models (at most any N parents). In situations where unrestricted ANBNs are used as classifiers, links between all variables and the target or class variable are forced. However, in learning the structure through the minimum description length, structure learning can be its own feature selection process, as links between target variables and independent variables are learned (through minimizing

MDL) rather than forced. This requires complete case data (i.e. no missing values), which is a significant limitation with clinical data.

Overall, learning an augmented Bayesian network allows for modeling of multiplicative effects, such as an increased likelihood of death if both cardiac output and mean pulmonary arterial pressure are low in a patient, versus high mean pulmonary arterial pressure and high cardiac output, which is more typical in early stages of the disease.(10) Lastly, the usability and accuracy of classic statistical methods tools are called into question when specific measurements are missing from patients and cannot be reasonably imputed.(73) This is a common problem with clinical data where a clinician's time and resources can be very limited. Whereas traditional PAH risk scores do not have a clear means of handling missing data, BNs can substitute the most likely variable given all others. A further advantage of BNs is their representation with a graphical structure. This increases user confidence as the model is not a "black box" and interactions between variables, the outcome, and the strength of the relationship can be visualized. An example of a TAN Bayesian network is shown earlier in the PHORA model, Figure 1.5.

Greedy Thick Thinning (GTT) is a means of both structure development and feature learning for a Bayesian network.(74) By using a three-phase sequence (i.e. “thickening” and “thinning” phase, proceeded by a “drafting” phase) to learn relationships between variables, in which the “thickening” phase establishes connections to maximize the joint probability of the structure, and the “thinning” phase removes connections to arrive at a minimal independence map. GTT is thought to approximate causal inference. Variables that might be correlated through a third, causal variable are not connected directly in a GTT structure, but rather connected through the third causal variable. This is depicted in

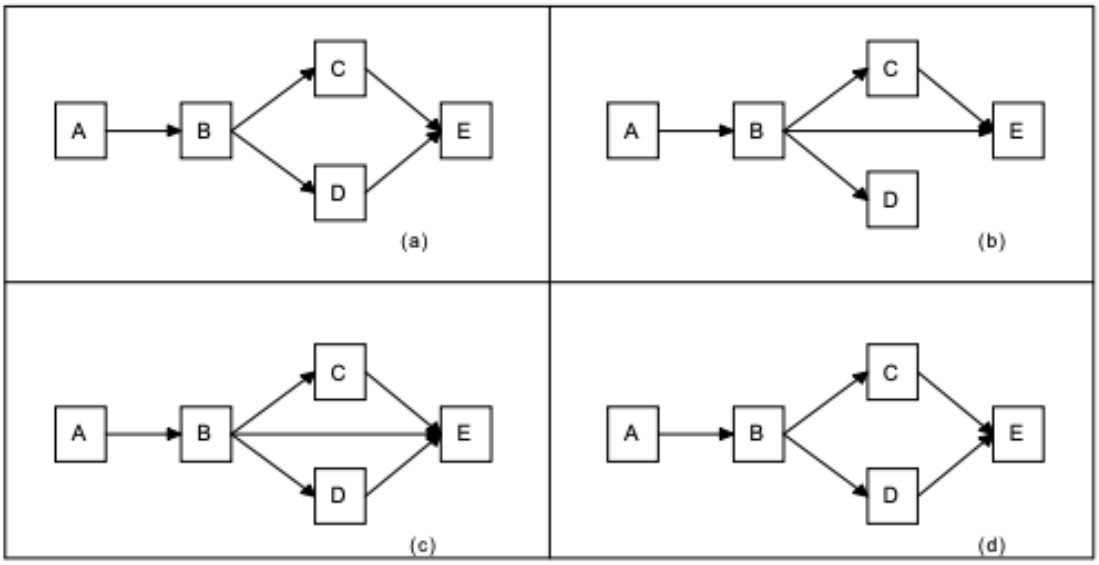


Figure 1.8.

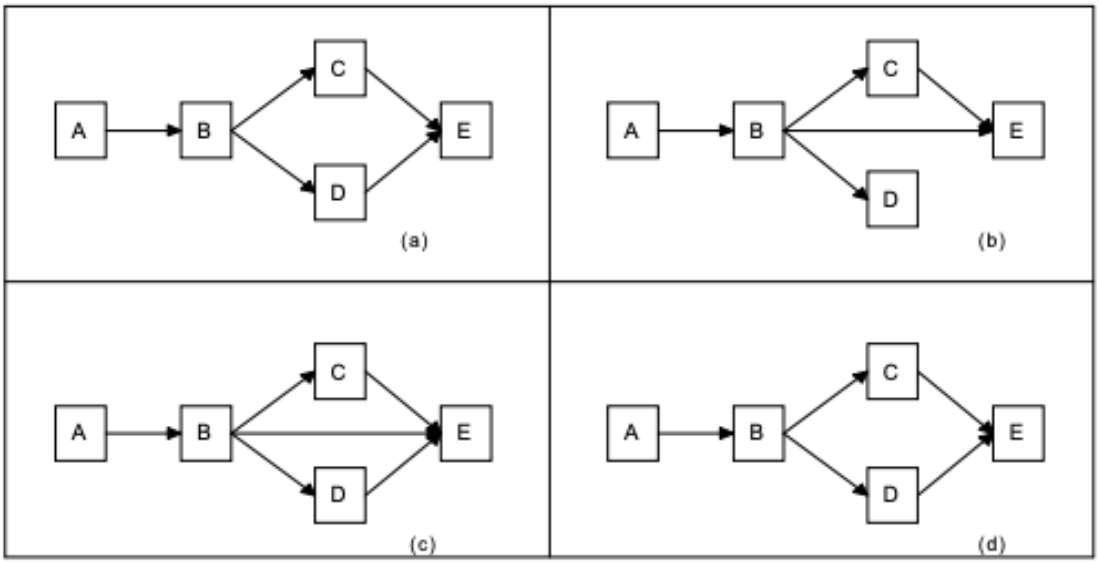


Figure 1.8. Greedy Thick Thinning Structure Learning. Reproduced with permission of authors [74]

In

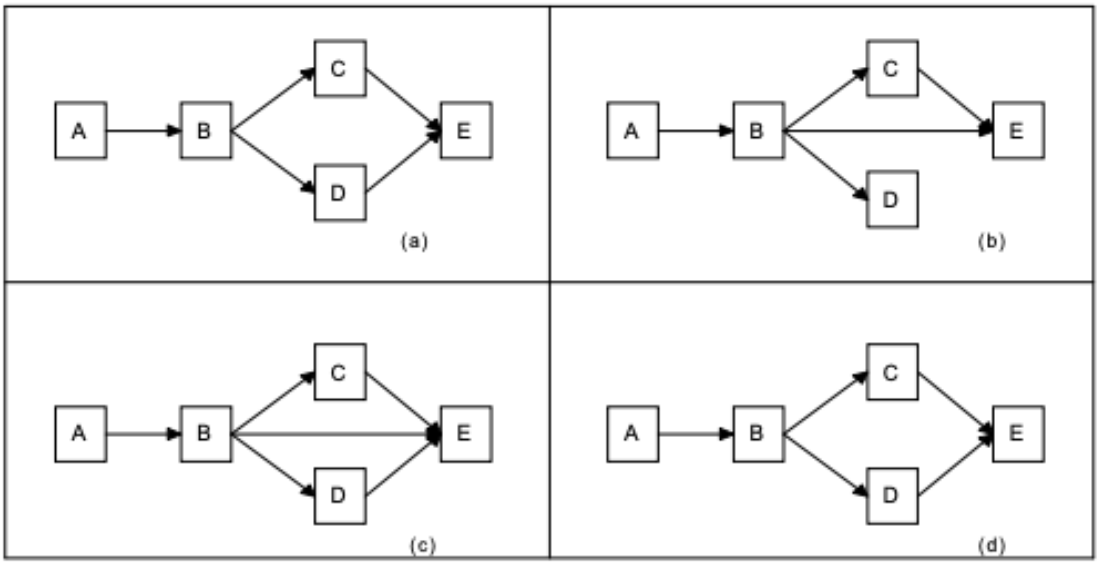


Figure 1.8(a), the true causal map is shown.

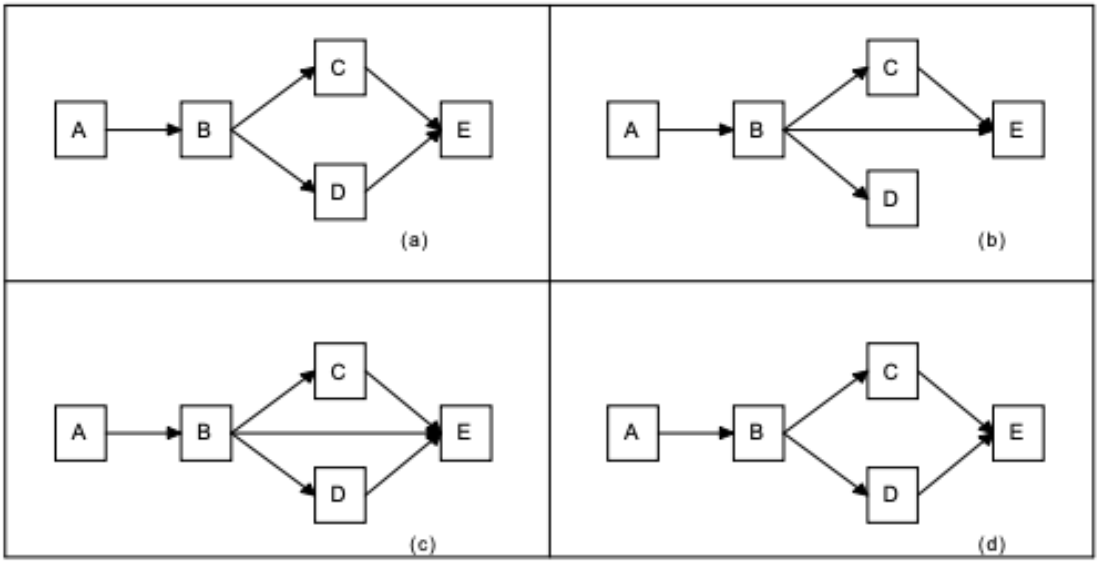


Figure 1.8(b) shows the graph that would be derived from the Chow-Liu algorithm, based on the maximum spanning tree, solved during the “drafting” phase of GTT. In this phase, node B and E are strongly correlated, and therefore linked, but this first phase cannot determine that nodes B and E are linked causally through node D.

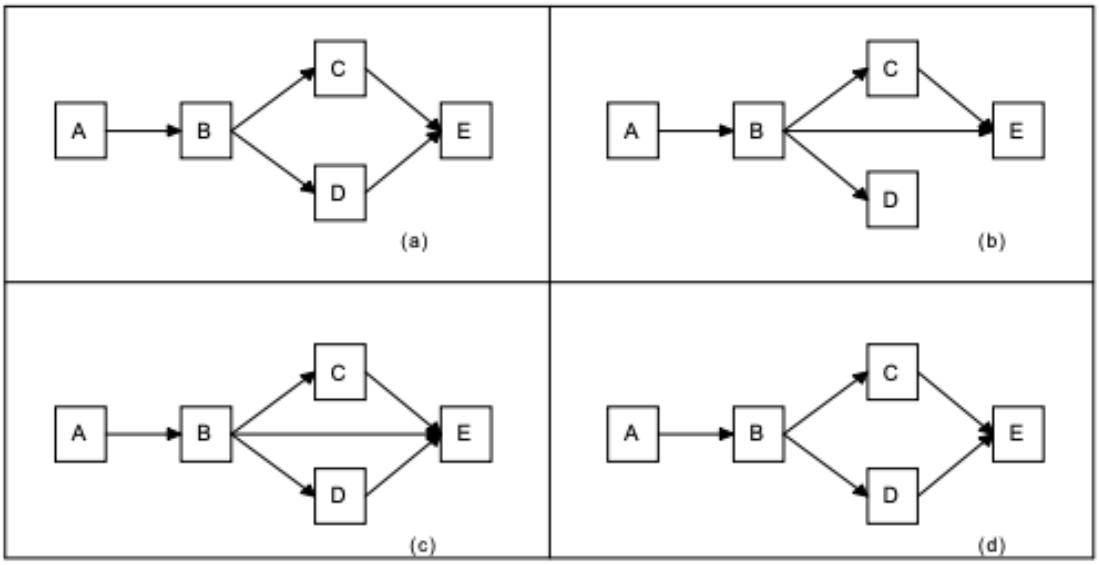


Figure 1.8(c) shows the fully connected graph, found during the “thickening” phase of GTT, as node E is still dependent on node D given node B.

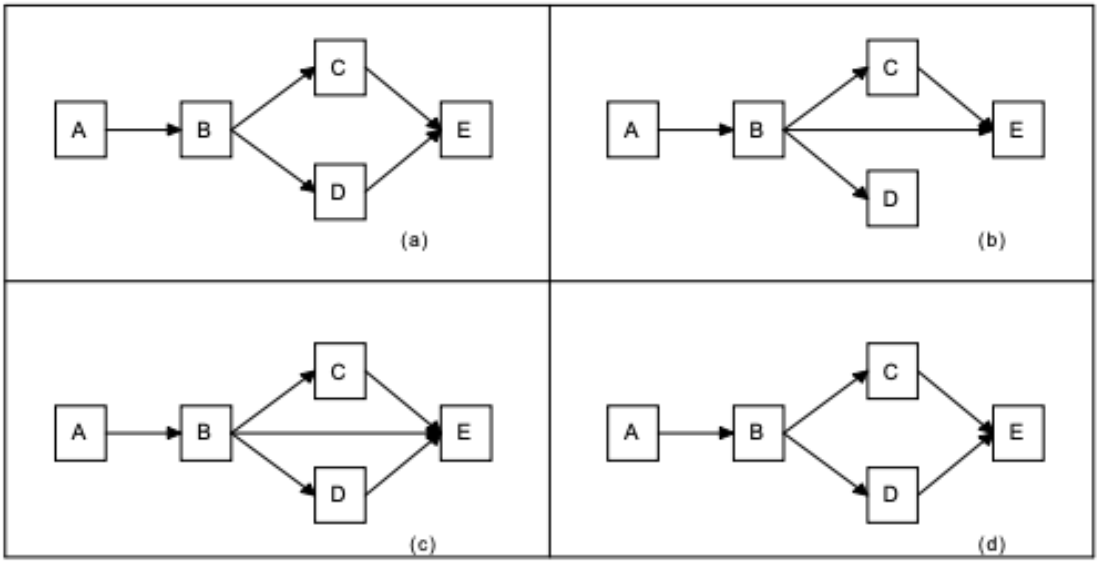


Figure 1.8(d) shows the final graph derived after the “thinning” phase of GTT, which determines that node E is not dependent on node B given node D. Through this three-step process, the true causal map is determined. This structure can then be used to determine which variables have direct influence on a target variable (i.e. one-year survival), by searching the structure for the target’s Markov blanket. The Markov blanket of a variable, within a graphical structure, are the variable’s parent nodes (nodes pointing to the variable), child nodes (nodes pointing away from the variable), and the variable’s co-parents (nodes that also point to the variable’s child s).(75) For example,

the Markov blanket of Node C, as shown in

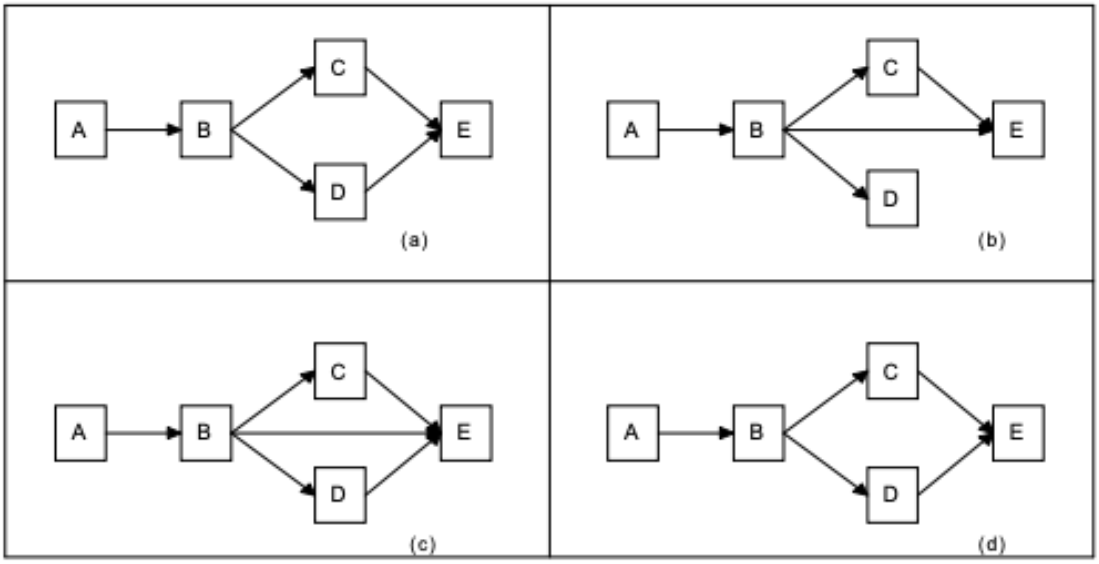


Figure 1.8, is Node B (its parent), Node E (its child), and Node D (its coparent). The target is conditionally independent of all other variables given its Markov blanket.

GTT structures are limited in that they must learn from complete case (no missing) data. This is a significant limitation especially for medical data, where data is both financially costly to collect and invasive to access, so missingness is very common. Therefore, while GTT structures have been shown to learn perfect mappings of true, underlying structures, they poorly approximate structures in situations where sample sizes are limited by data missingness.

The PC algorithm (named after its authors, Peter Spirtes and Clark Glymour) can also be used to learn a structure for a Bayesian network or for feature learning.(76) The algorithm depends on a series of statistical independence tests to determine structure connectivity. After the structure is learned, the Markov blanket of the target variable can

be located to identify the features that are most relevant. Again, however, this algorithm depends on complete case data.

Novel Methods of Feature Selection and Learning for Bayesian Networks

Development of any machine learning model is a question of efficient search optimization methods. This includes efficient search of “learned” model parameters and “tuned” hyperparameters. In the *Bayesian Network Modeling* section, the means of learning model parameters as well as tuning network structure has been discussed. Further, examples are given for how the structure learning process itself can be used to learn features. However, there are techniques external to the structure learning process that can be used to preprocess and choose the right features for the network. In this regard, feature selection and preprocessing are hyperparameters that can also be tuned to improve model performance.

The following discussed methods are agnostic in that they can be generalized to any (discrete) machine learning model and they will be explored in the context of creating an improved prognostic tool for PAH. Specifically, these methods will be discussed in the framework of optimizing a (non-naïve) Bayesian network that depends on discrete features to predict a binary outcome, as such was the intended application for this dissertation. All of the following methods are applied in Chapter 4.

Differential Evolution

Differential evolution (DE) is a method for efficient global search given a specific minimization task for non-convex objective functions.(77) Its goal is to obtain or approximate a solution that ideally globally minimizes the objective function. It is more

robust against getting “stuck” at local minima compared to gradient-based or other stepwise search methods.

In essence, differential evolution accomplishes the minimization task by generating an initial random population of candidate solutions, combining candidates into a new candidate through “mutation”, and then mixing parameters to move towards a trial vector that is the current best guess, referred to as “crossover”. This is repeated for a set number of iterations, determined by the user, or until the solution does not change significantly over the next iteration, a threshold that can also set by the user.

The initial population can be set to any size, but it is best to choose a size that adequately represents the full parameter space. Specific methods have been developed solely for the purpose of choosing a well-represented initial population, such as Latin Hypercube sampling, in which the coverage of the available parameter space is maximized by only sampling once from a designated “hypercube” area. This is depicted in Figure 1.9 (b), where each “x” designates a sample in the parameter space of *A* and *B*, but does not sample from a value of “X_1” twice nor a value of “X_2” twice. This is contrasted with purely random sampling as depicted in Figure 1.9 (a), where two samples have very different “X_1” values, but extremely similar “X_2” values, by pure chance. A method that is slightly more restrictive of Latin Hypercube is orthogonal sampling, as depicted in Figure 1.9 (c), where each parameter space is partitioned into areas of equal probability and sampled simultaneously such that each subspace is sampled with the same density. While orthogonal sampling slightly improves upon ensuring the full variability of the parameter space is covered, it is also more computationally expensive.(78)

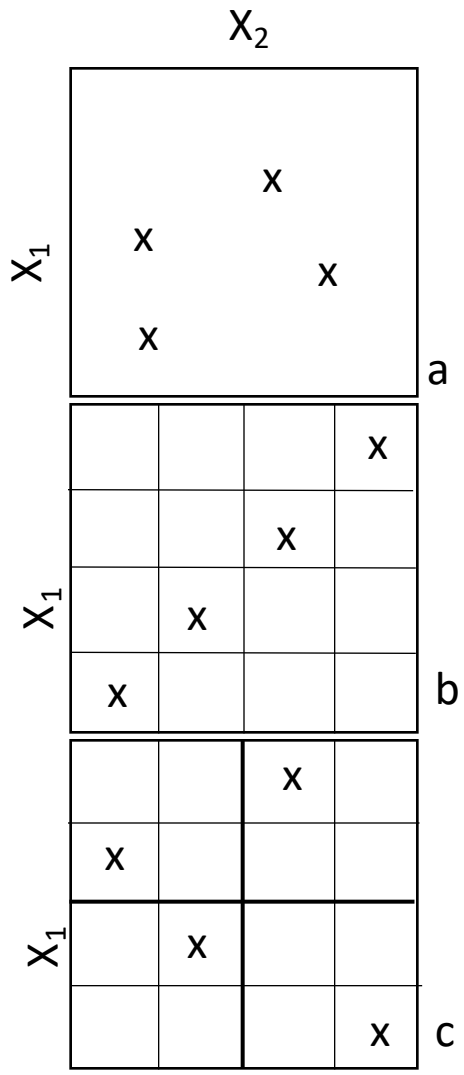


Figure 1.9: Illustration of Random Sampling, Latin Hypercube Sampling, and Orthogonal Sampling.

Next, mutation is achieved through taking the weighted difference between two randomly chosen population vectors and adding this difference to a third randomly chosen vector in the population. The weight of that difference can be specified by the mutation constant. The mutation constant can be chosen at random at mutation time, in a process known as “dithering”. This allows for faster convergence to a solution.

Generally, there are three methods for applying differential evolution for feature selection for machine learning: 1) “filter” (i.e. unsupervised) based methods, where the goal is to minimize the number of features without knowing the predicted target or label, by choosing the group of features that maximizes some underlying representation of the data, 2) “wrapper” based methods, where the label and model are known, generating multiple candidate models with different feature sets and evolving them towards a globally optimized model with a unique feature set, 3) hybrid “filter-wrapper” methods, where the predictive power of each feature is known, and the goal is to maximize estimated predictive power and the underlying representation of the data with as few features as possible.(79; 80) Computationally, wrapper-based methods are very expensive, as they require model generation and parameter learning for each candidate in the population. There’s also a greater propensity to overfit the model on training data, unless caution is used to ensure that performance for each candidate model represents generalized performance (such as cross-validated performance). Even then, as cross-validated performance tends to overestimate test performance, overfitting is still possible. While less computationally expensive, filter-based methods can bias models towards feature that represent much of the underlying data pattern while neglecting features that might significantly increase performance when taken into account with all other features, but represent a small degree of the overall variance. Hybrid based methods, therefore, constitute a sort of “best of both worlds”, though by not taking into account the structure of the model, may not guarantee that predictive power is fully realized.

Decision Trees

Decision trees are, of themselves, machine learning models that learn a series of if-else statements to make a final prediction on a label. However, they've increasingly been used as a means of both feature selection and feature learning discretization.(81) While there are many versions of decision trees, specific focus is given to Classification and Regression Trees (CART), as they are capable of learning if-else statements for mixed data type features (both continuous and discrete).(82)

Decision trees can learn features by identifying an if-else statement for the feature that can reduce the Shannon entropy (shown in Equation 5) of the class label data, when that class is discrete or binary.

$$H(X) = -\sum_{i=1}^{n} P(x_i)\, log_2\, P(x_i), \quad (5)$$

where $H(X)$ is the Shannon entropy of the vector X (measured in bits of information), X is the vector of class observations for n classes = $\langle x_1, \ldots, x_n \rangle$, $P(x_i)$ is the probability of the observation of the class.(82) With a binary class label, when the probability of either class is perfectly split 50/50, Shannon entropy is maximized (1 bit of information). Entropy of the class label is reduced when a decision tree learns how to increase the probability of one label by sorting datapoints on a feature or set of features. For example, assume a dataset is shown where 50% of all patients have a diabetes diagnosis. However, if you selected patients only with a fasted blood glucose level greater than 125 mg/dL, perhaps 80% of those patients have a diabetes diagnosis. Thereby, the entropy of the data set is reduced (0.72 bits), as long the fasted blood glucose level of each patient is known. The decision tree has learned that its useful to know the fasted blood glucose level when making a diabetes diagnosis.

By reducing entropy, the class labels come closer to being “sorted”, allowing the decision tree to create an approximate mapping of features to the observed probability of a specific label in training data. If a decision tree is able to learn a perfect classification of the data, it has ensured that the probability of observing a label given specific if-else conditions is 100% (i.e. the entropy of the training data partitioned is now zero). If a feature cannot reduce the entropy of the class label, or cannot reduce it a reproducible way (i.e. consistent cross-validation performance), it can be deemed to be insufficiently predictive of the label and discarded.

Alternatively, decision trees can also learn if-else statements by optimizing other scoring systems, such as the negative Brier score, show in Equation 6.

$$Negative\ Brier\ Score = -\frac{1}{N}\sum_{i=1}^{N}\big(f(i) - o(i)\big)^2 \text{ , (6)}$$

where N is total number of forecasted events, f(i) is the forecasted probability of event *i*, o(*i*) is the true outcome of *i* (either *i* = 1 if the event did occur, or *i* = 0 if the event did not occur). The use of this scoring system is preferred when the calibrated accuracy of the model matters more than the dichotomized prediction. This is true especially when modeling clinical outcomes such as survival, as determining if a patient is at a “high risk” for a mortality event is typically more relevant than attempting to determine a nondeterministic (i.e. somewhat random) outcome. Other decision trees can maximize the significance of statistical tests, which is the task its applied to with “survival trees”, a technique applied in Chapter 1.(83) These decision trees maximize the estimated survival difference between populations by identifying a “cut-point” for a variable or score, based on statistical significance testing. A cut-point is the value of the inequality statement learned by a decision tree.

Decision trees are unique from many commonly used forms of traditional statistical methods or machine learning models because they are capable of learning non-linear relationships between the label and the features. For example, as earlier stated, if patients with both a low value of a predictor (such as low BMI, or underweight status) and a high value of that same predictor (such as high BMI, or obese status), both conferred a higher risk of death, while having a normal to slightly overweight BMI conferred a reduced risk of death, a decision tree could learn this phenomenon.

A major weakness of decision trees is that they can tend to learn spurious cut-points that, by chance, reduce the entropy of the label, but are not reproducible. This is another example of model overfitting. For this reason, controlling the number of cut-points learned can reduce the chances of learning spurious relationships. This can be achieved by setting certain restrictions on the decision tree, such as the maximum depth (i.e. number of splits) it can achieve, or enforcing that splitting cannot continue if the number of samples left to split on is low. Further, decision trees can be "pruned", in that a cut-point learned can be tested, typically with a statistical test, to determine if the cut-point identifies truly significantly different groups. If the significance test is not satisfied, the cut-point is "pruned" from the tree and the two groups are re-merged.(84)

Physiological Modeling of Right Heart Failure

As proposed by Sarma and Reinersten, physiology should be the "lingua franca" (i.e. common language) of both machine learning decision support systems and clinicians.(8) Indeed, most graphical structure models imitate influence diagrams used to show relationships between different biomarkers. Both the careful design of features

(known in the machine learning scientific community as “feature engineering”) and the design of a model itself can be driven by physiological understanding of a disease. Therefore, improved physiological understanding of right ventricular failure is beneficial, in and of itself, for improving treatment strategies for PAH, and should also lead to better machine learning models to further optimize treatment strategies or predict patient outcomes.

Specifically, right ventricular energetics has been poorly studied in the setting of pulmonary arterial hypertension. Although clinicians recognize decreases in cardiac output as a clear sign of disease progression, there is little discussion of how this creates a negative feedback loop by which reduced flow to the right coronary artery prevents the right ventricle from producing the power needed to overcome a high mean pulmonary arterial pressure. This would result in progressively lower cardiac output and further worsening of the patient’s condition. While there are autonomic compensatory mechanisms to combat this, including dilation of the right coronary artery, myocardial angiogenesis, etc., understanding the total energetic strain on the right ventricle could provide greater clinical insight as to which patients are most vulnerable to acute stressors (infection, physical exertion, coronary artery disease, or other comorbidities). Hence, the most common, Suga-Sugawa model of ventricular energetics is reviewed, and its limitations are discussed.

Suga Sugawa model of Ventricular Energetics

The study of myocardial energetic needs is considered crucial to the treatment of left ventricular heart failure. Current clinical practice employs the use of the Suga-Sugawa

Pressure Volume Area (PVA) model to estimate total ventricular work demands from pressure-volume loops.(85) The PVA model of myocardial energetics is based on the assumption that there are two main types of work performed by the ventricle: 1) stroke work (W) during the ejection phase, 2) work that is stored as "potential energy" (U) from myocardial stretching during diastole (shown in **Error! Reference source not found.**).(85) These two forms of work compose the entire pressure-volume area.

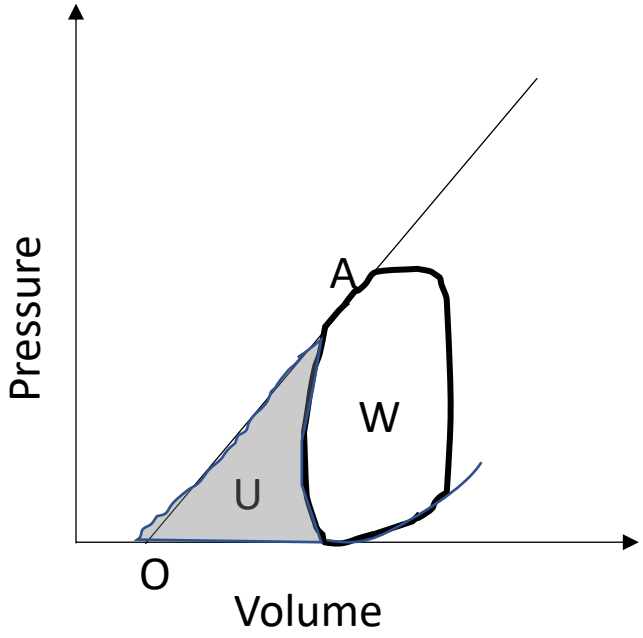


Figure 1.10. Suga-Sugawa Model of Ventricular Energetics.

Stroke work is calculated as the total area under the pressure-volume loop starting from the opening of the tricuspid valve, diastolic filling of the right ventricle, to tricuspid valve closure, isovolumic contraction, pulmonic valve opening, right ventricular ejection, pulmonic valve closing, and right ventricular relaxation. The "work" conducted during the diastolic phase is calculated as the remaining area under a (typically) linear end-systolic

pressure volume relationship (ESPVR) curve, shown in Figure 1.10 as the line from point O to point A. This potential energy can defined mathematically by Equation 7.(85)

$$Potential\ Energy = \int_{V_0}^{EDV} P(v), \quad (7)$$

where *EDV* is the ventricular end-diastolic volume, V_0 is a hypothetical "dead" volume in the ventricle, defined by the x-intercept of the ESPVR curve, *P(v)* is the instantaneous pressure at volume *v* of the ESPVR curve. Multiple pressure-volume loops with different preload amounts are required to calculate the ESPVR curve and the corresponding area, which is achieved through manual venous manipulation. This method is considered too invasive to be considered for routine clinical practice. There is also some debate as to whether the ESPVR curve should be linear or nonlinear, though the differences in calculation of potential energy are typically negligible. Single-beat methods with a single preload have been proposed but are considered less accurate.(86)

Overall, while the pressure-volume area is strongly correlated with myocardial oxygen consumption, it reflects only 40% of total oxygen consumed (i.e. the empirical slope is smaller than it should be, by roughly a factor of 2).(87) Further limitations of the Suga-Sugawa method are that it does not follow physiological or thermodynamic principles. In addition, the determination of V_0 (as shown in Equation 6) can be negative and does not reflect any sort of physiological or anatomical purpose. Lastly, the method does not consider the energy required to raise the ventricular end-diastolic volume to the initial ejection pressure during systole, a form of "isovolumic work". This work more accurately portrays the stored "potential energy" conducted by the right ventricle, what *Elbeery et al* in 1995 described as an "internal index of heat".(87) Alas, the myth of the

PVA model of ventricular energetics has only very recently been dispelled as inconsistent with thermodynamics.(88)

The pressure-volume area of energetics, likely due to its inaccuracy, has not been shown in the literature to be a useful means of study right ventricular failure. Yet it's very likely that right ventricular energetics play a significant role in understanding right heart failure, including due to PAH. Therefore, the derivation of a more accurate model is required to further understand the role of RV energetics in PAH disease severity.

Summary of Study

This dissertation proposes a multi-pronged computational approach to improving clinical decision science specific to PAH, which could also be more widely applied to other multi-faceted rare diseases:

1) Application of risk algorithms for the enrichment of clinical trials through retrospective simulated optimization of selection of a high-risk patient cohort (Chapter 2)
2) Application of risk algorithms for the retrospective study of treatment optimization (Chapter 3)
3) Improvement of risk algorithms with bottom-up machine-learning development (Chapter 4)
4) Establishing new metrics of right ventricular function through novel energetic modeling (Chapter 5)

Chapter 2 : Risk Enriched Clinical Trial Design

Introduction

Pulmonary arterial hypertension (PAH) is a collection of rare and progressive disorders of the pulmonary vasculature with no known cure. Clinical trial endpoints used to demonstrate drug efficacy efficiently and effectively in PAH trials are either insufficient or inefficient at demonstrating clinical benefit.(89; 90) Early-era PAH trials were primarily required to show a statistically significant increase in six-minute walk distance (6MWD) for demonstrating drug efficacy. However, it is now known that improvement in 6MWD is only weakly associated with reductions in clinical events. Moreover, demonstrated improvements in 6MWD are typically small (average 30 m increase on an average baseline of 350 m), with debatable clinical relevance.(12) Contemporary PAH clinical trials switched their focus to complex determinants of therapeutic efficacy, such as time to clinical worsening, which is a composite endpoint of death, hospitalization, and other measures of disease progression. Although such event-driven endpoints demonstrate a benefit with clear clinical relevance to patients, clinical worsening is relatively infrequent. Hence, successful PAH trials end up requiring large-scale patient enrollment for lengthy durations, with substantial economic expenditure.(12)

Existing, validated risk-prediction algorithms derived from PAH registry data could be used to identify patients at intermediate- and high-risk of clinical worsening

using baseline clinical trial data. The goals of this chapter are to demonstrate that these algorithms are prognostic of a clinical worsening event and, in simulated scenarios, a patient cohort enriched with higher risk patients (as identified by risk algorithms) can demonstrate a significant treatment benefit with a substantially smaller sample size compared to other, more simplistic measures such as six-minute walk distance.

Methods

Three contemporary PAH risk-prediction algorithms were used to stratify patient risk of clinical worsening at baseline: Comparative, Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension (COMPERA)(59), the French pulmonary hypertension registry score (French)(60), and the United States Registry to Evaluate Early and Long-Term PAH Disease Management (REVEAL 2.0)(61). A description of the design of all these scores, including inclusion and exclusion criteria, is provided in Chapter 1. Three contemporary PAH trials, A Study of First-Line Ambrisentan and Tadalafil Combination Therapy in Subjects With Pulmonary Arterial Hypertension (AMBITION)(42), Selexipag in Pulmonary Arterial Hypertension (GRIPHON)(37), and Study of Macitentan on Morbidity and Mortality in Patients With Symptomatic Pulmonary Arterial Hypertension (SERAPHIN)(16) were chosen for analysis since they measured time to clinical worsening as their primary endpoint. Definitions of clinical worsening for each trial are provided in Table 2.1.

Clinical Trial	Definition of Clinical Worsening
AMBITION	Any one or more of the following events: • All-cause death • Hospitalization for worsening PAH (includes lung, heart-lung transplant, atrial septostomy, initiation of parenteral prostanoid therapy) • Decrease of more than 15% from baseline in 6-minute walk distance (6MWD) combined with WHO functional class (FC) III or IV symptoms at two consecutive visits separated by at least 14 days • Any decrease from baseline in 6MWD separated by at least 14 days, WHO FC III symptoms assessed at two visits separated by at least 6 months
SERAPHIN	Any one or more of the following events: • All-cause death • Initiation of parenteral prostanoid therapy • Lung transplantation • Atrial septostomy • Decrease of at least 15% from baseline in 6MWD at two visits within 2 weeks, combined with worsening symptoms (change from baseline of WHO FC, or no improvement for FC IV patients at baseline, or the appearance or worsening of right heart failure symptoms that did not improve with oral diuretic treatment), and the need for additional treatment for PAH
GRIPHON	Any one or more of the following events: • All-cause death • Hospitalization for worsening PAH • Initiation of parenteral prostanoid therapy or long-term oxygen therapy • Need for lung transplantation or balloon atrial septostomy • Decrease of at least 15% from baseline in 6MWD at two visits on different days, combined with change from baseline of WHO FC for FC II or III patients at baseline or the need for additional treatment for PAH for FC III or IV patients at baseline

Table 2.1: Definitions of Clinical Worsening in Pulmonary Arterial Hypertension Clinical Trials

Data across trials were unified and standardized, in terms of their units for clinical measures and demographic encoding, for testing the investigational risk algorithms. The variables used in each algorithm and cut-points used to determine risk are shown in Chapter 1, Table 1.4.

All risk algorithms were applied as intended in the original publication except for COMPERA, which was left as a continuous variable and not rounded to the nearest integer after averaging. The following variables were unavailable across all clinical trials: diffusing lung capacity and hospitalization six months prior to randomization for REVEAL 2.0 and mixed venous oxygen saturation for COMPERA. Estimated glomerular filtration rate for REVEAL 2.0 was calculated using the *Levey et al* equation, which considers race, age, and sex.(91)

For statistical analysis, receiver-operating characteristic (ROC) curves were generated for each algorithm to determine their ability to predict clinical worsening as defined by the trial's original primary endpoint. Briefly, an ROC curve provides a quantitative assessment of overall predictive performance by plotting the value of sensitivity and specificity at each potential threshold of the risk score. Highly predictive scores will have an area under the ROC curve near one (1), poorly predictive scores will have an area under the ROC curve near one-half (0.5), which indicates that the score is no better than guessing at random.(92) Algorithms were benchmarked against a traditional clinical means of patient risk stratification (New York Heart Association [NYHA] Functional Class used in isolation) via nonparametric statistical analysis (i.e., bootstrapping) to determine statistical significance of the difference in the AUCs. The predictive value of each algorithm was compared to univariate biomarkers (6MWD, NT-

proBNP, mean right atrial pressure, cardiac index, and pulmonary vascular resistance) to determine if they could provide the same degree of predictive performance as a multi-variable risk score. Patients who were censored early from the primary endpoint were imputed as event-free and a sensitivity analysis was conducted to determine the impact of this assumption. To avoid confounding baseline risk and treatment effects, only the placebo populations in each trial were used for ROC analysis.

As both COMPERA and REVEAL 2.0 generate risk scores on a near continuous scale, a consistent means was required to define cut points that allowed for simplified patient groupings for enrichment: low- versus intermediate- versus high-risk, with the possibility for very-low and/or very-high depending on algorithm precision. Further, cut points were optimized such that the high-risk patients not only had more clinical worsening events, but also a faster time to clinical worsening. To that end, a survival tree analysis was applied to the pooled placebo population to determine such cut points for each algorithm (applied via rpart R package, v4.1-15, 2019).(83; 93) Survival trees identify cutpoints for each algorithm by identifying groups with a statistically significant difference in time to clinical worsening. The optimal number of cutpoints is determined by multiple rounds of data splits and tests to ensure each group has a statistically significant difference in survival. Each survival tree was optimized to find the largest number of cut points such that each group had a statistically significantly different survival curve (per exponential regression) using a 10-fold cross-validation. Using the largest number of cut points allows for identification of an optimally high-risk patient group from each algorithm that could be recommended for prognostic enrichment without a trial sponsor conducting their own analysis.

Identified cut points were applied to the pooled treatment population to determine whether each simplified risk group saw a benefit in treatment, as determined by Cox Proportional Hazards. This analysis was necessary to support the goals of prognostic enrichment as a means to establish drug efficacy in high risk groups but still prescribe later in lower risk patient populations. $P < 0.05$ was used to determine a statistical significance. Incidence rate of clinical worsening for each group and each treatment arm was calculated as events per 100 patient-years.

Sample size estimates to demonstrate drug efficacy in each clinical trial were recalculated by resampling from patients with no missing algorithm data and employing the method originally proposed by Freedman in 1982, which uses pilot data to estimate probability of clinically worsening event in each treatment arm (applied via the powerSurvEpi R package, v 0.1.0, 2018) and then calculates final sample size needed based on expected value and variance of event rates.(94; 95) Resampling was conducted to reflect multiple enrichment strategies: 1) selecting only intermediate- and high-risk patients (intermediate-high strategy); 2) selecting 50% of patients from high-risk only and 50% from all other risk groups (high-other strategy); and 3) selecting 100% of patients from high-risk only (high-risk strategy). A nonparametric bootstrap analysis was used to generate 95% confidence intervals for each estimated sample size determined from an event-driven power analysis. The parameters of the power analysis, namely confidence level, anticipated effect size (i.e. hazard ratio of clinical worsening, where smaller values indicate greater treatment benefit), and power to detect the hypothesized treatment effect were kept as published in the original trial's statistical

design (see Table 2.2). Discussion of the implications of differences in statistical design are included in the Limitations section.

Trial	Power	Confidence Level	Estimated Effect Size (Hazard Ratio)
AMBITION	97%	5%	0.47
GRIPHON	90%	1%	0.65
SERAPHIN	90%	0.5%	0.55

Table 2.2: Statistical Analysis Design Per Original Clinical Trial

The procedure for the nonparametric method of sample size estimation was as follows:

1) Identify the patient population in the original trial that meets the risk level required.

2) Using bootstrapping, resample an equal number of patients from both the placebo and treatment arm to simulate pilot data.

3) Using the powerSurvEpi function “ssizeCT” to calculate the estimated sample size using the established pilot data and parameters as defined by the clinical trial’s original statistical plan (see Table 2.2).

4) Repeat this procedure 1000 times to capture variance in estimates due to random resampling, providing a 95% confidence interval and the average treatment time of patients.

While no attrition rate is specified in this analysis, resampling from the original trial populations allowed results to reflect sample size estimations in the presence of early withdrawals. Mean times in trial (i.e., treatment time) were also calculated for each enrichment strategy per algorithm. Sample size and time in trial from simulations are presented as a percent reduction from simulations using the non-enriched

subpopulation, where higher reduction indicates lower sample size, shorter treatment time, and therefore improved trial efficiency.

Clinical trial enrichment can potentially enable smaller clinical trials and thus cost savings. However, the patient screening necessary for trial enrichment can be burdensome and costly. Thus, the likelihood of finding a patient fitting each algorithm's risk groupings was determined by calculating the ratio of total patients screened to total patients enrolled per risk category in the pooled trial dataset. Ratios are presented as the number of patients that must be screened to enroll 100 patients, where a higher ratio is indicative of higher screening numbers, lesser screening efficiency, and greater cost.

Finally, a hypothetical cost-savings exercise was conducted with the GRIPHON trial to demonstrate the benefit of enrichment in balancing the increased cost of screening by reducing the cost of treating enrolled patients (i.e., research costs) and to determine the optimal enrichment strategy for balancing these costs. Minimum estimated costs per patient were based on figures reported by Ryan and colleagues and other clinical procedure price estimations.(89),(96),[97] Costs of treatment are provided in Table 2.3.

Study Element	Baseline Cost ($ USD)	Iterations per Study Per ITT Patient	REVEAL 2.0 Screening	COMPERA Screening	French Screening	Non-Enrichment Screening
Informed Consent Processing	150	0	1	1	1	1
History and Physical Examination	500	0+	1	1	1	1
Vital sign assessment	50	3+	1	1	1	1
Right Heart Catheterization	3500	0+	1	1	1	1
6-Minute Walk Test	550	3+	1	1	1	1
NT-proBNP	140	0+	1	1	-	-
Lung Capacity Test	500	-	1	-	-	-
Mixed Venous O_2 Saturation	200	-	1	1	-	-
Creatinine	50	0+	1	-	-	-
IRB Fees	4,000	1	-	-	-	-
Study Drugs	12,100	1	-	-	-	-
Total		$13,900+ per treated patient; $1,800+ per placebo patient	$5,640 per patient	$5,090 per patient	$4,750 per patient	$4,750 per patient

Table 2.3: Cost Analysis per Pulmonary Arterial Hypertension Clinical Trial

.

The mean time to clinical worsening or censor, as calculated from power simulations, was used in the cost analysis as the average treatment time per patient receiving the study drug. For this analysis, it's assumed that every patient, regardless of risk algorithm, required a right heart catheterization procedure in screening to confirm PAH Group I diagnosis. Although this study used fewer variables to stratify patients, final estimated screening cost for each algorithm reflected the cost of collecting all data required for each algorithm. To account for the worst-case of screened patients failing due to selection criteria outside of the matching risk level (such as very low 6MWD or prostacyclin analogue background therapy), the number of patients required for screening was calculated using the following equation:

$$Patients\ Needed\ for\ Screening = Patients\ Needed\ for\ Enrollment \times (Screen\text{-}to\text{-}Enrollment\ Ratio\ of\ GRIPHON) \times (Screen\text{-}to\text{-}Enrollment\ Ratio\ for\ Enriched\ Population), \quad (8)$$

Percent cost saving was calculated as the percent difference between total cost for the given enrichment strategy and the estimated cost with no enrichment.

Results

From a total of 1,769 patients, N=976 (55%) and N=793 (45%) patients were identified with all required variables from the treatment and placebo groups, respectively. As shown in Figure 2.1 all algorithms performed similarly in their ability to predict clinical worsening (COMPERA AUC 0.70, 95% CI: 0.66-0.73; French Score AUC 0.66, 95% CI: 0.63-0.70, REVEAL 2.0 AUC 0.70, 95% CI 0.66-0.73).

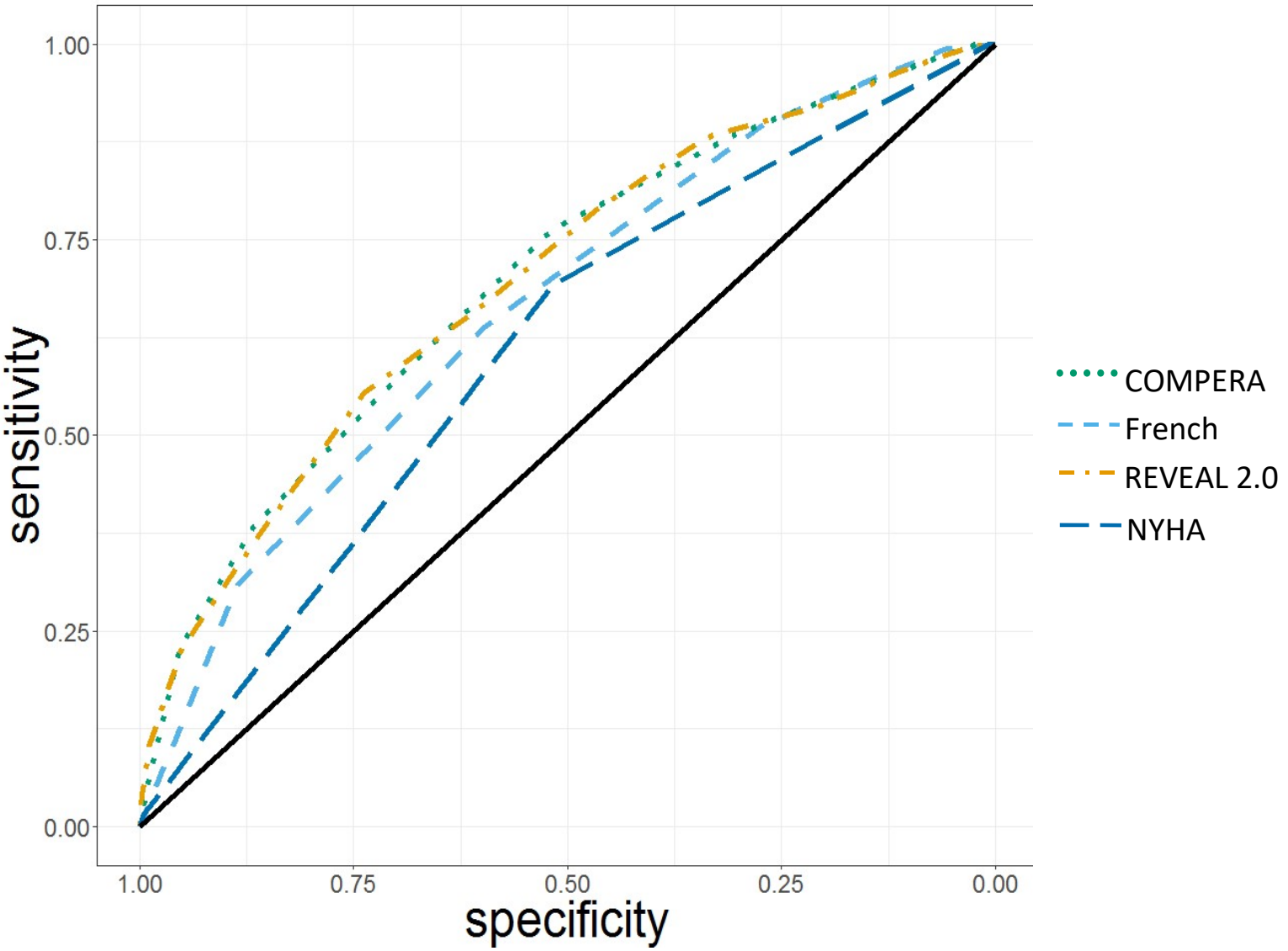


Figure 2.1: Receiver-Operating Curves (ROCs) for Each Investigational Algorithm.

Each investigational algorithm outperformed NYHA functional class (AUC 0.61, 95% CI: 0.57-0.64) at predicting clinical worsening prognosis (COMPERA p-value = 2.26×10^{-6}, French p-value = 6.5×10^{-4}, REVEAL p-value = 1.63×10^{-6}). Further, all risk algorithms had better performance in absolute AUC value for predicting clinical worsening than singular clinical variables (Table 2.4).

Clinical Variable	**ROC-AUC [Confidence Interval]**
Six Minute Walk Distance	0.65 [0.61-0.69]
NTproBNP	0.65 [0.61-0.69]
Mean Right Atrial Pressure	0.62 [0.58-0.66]
Cardiac Index	0.60 [0.56-0.64]
Pulmonary Vascular Resistance	0.63 [0.59-0.67]

Table 2.4. Prediction of Clinical Trial Endpoint by Single Clinical Variables in Isolation.

In sensitivity analyses to test the assumption that early censored patients were event-free, all algorithms perform worse if patients with a censor time of less than 3 years are removed from analysis, though excluding patients censored before one year largely did not change the AUC of each algorithm (see Appendix).

It's worth noting that REVEAL 2.0 was, however, the most robust against this assumption and largely maintained its prognostic performance.

Application of survival tree analysis to identify risk groups

Each investigational algorithm identified at least three unique risk groups with statistically significantly different time to clinical worsening rates, using survival tree analysis. Cut-points and incidence rates, as number of events per 100-patient-years, for each risk group are shown in Table 2.5.

Algorithm	Risk Group	Pooled Placebo Incidence Rate of Clinical Worsening (Events per 100-pt-yrs)	Pooled Treatment Incidence Rate of Clinical Worsening (Events per 100-pt-yrs)	Hazard Ratio for Reduction of Clinical Worsening Rate [95% Confidence Interval]
French	Low (2-3)	17.10	10.47	0.61 [0.46-0.81]
	Intermediate (1)	29.75	18.19	0.60 [0.46-0.79]
	High (0)	51.44	25.08	0.50 [0.38-0.67]
COMPERA	Low (≤ 1.7)	13.90	8.27	0.59 [0.42-0.85]
	Intermediate (>1.7-2.1)	27.37	16.36	0.59 [0.46-0.76]
	High (> 2.1)	52.52	24.81	0.48 [0.37-0.62]
REVEAL 2.0	Very Low (≤ 5)	11.00	6.61	0.60 [0.36-0.99]
	Low (6-8)	21.19	11.97	0.56 [0.42-0.74]
	Intermediate (9-10)	36.90	23.07	0.63 [0.48-0.83]
	High (>10)	72.32	34.52	0.48 [0.34-0.67]
NYHA	Low (≤ II)	16.69	9.79	0.58 [0.43-0.78]
	Intermediate (> II)	35.02	20.95	0.60 [0.50-0.73]

Table 2.5: Incidence Rate and Treatment Effect per Risk Group as Defined by Different Risk Algorithms

REVEAL 2.0 was the most precise and identified four statistically significantly different ranked groups for clinical worsening ($p < 2\times10^{-16}$; its full survival tree is shown as an example in Figure 2.2), specifically identifying an additional very low-risk group and its high-risk group had a much higher incidence rate than COMPERA's or French score's. Only three different groups were identified using either COMPERA ($p < 2\times10^{-16}$) or the French score ($p = 8.98\times10^{-16}$).

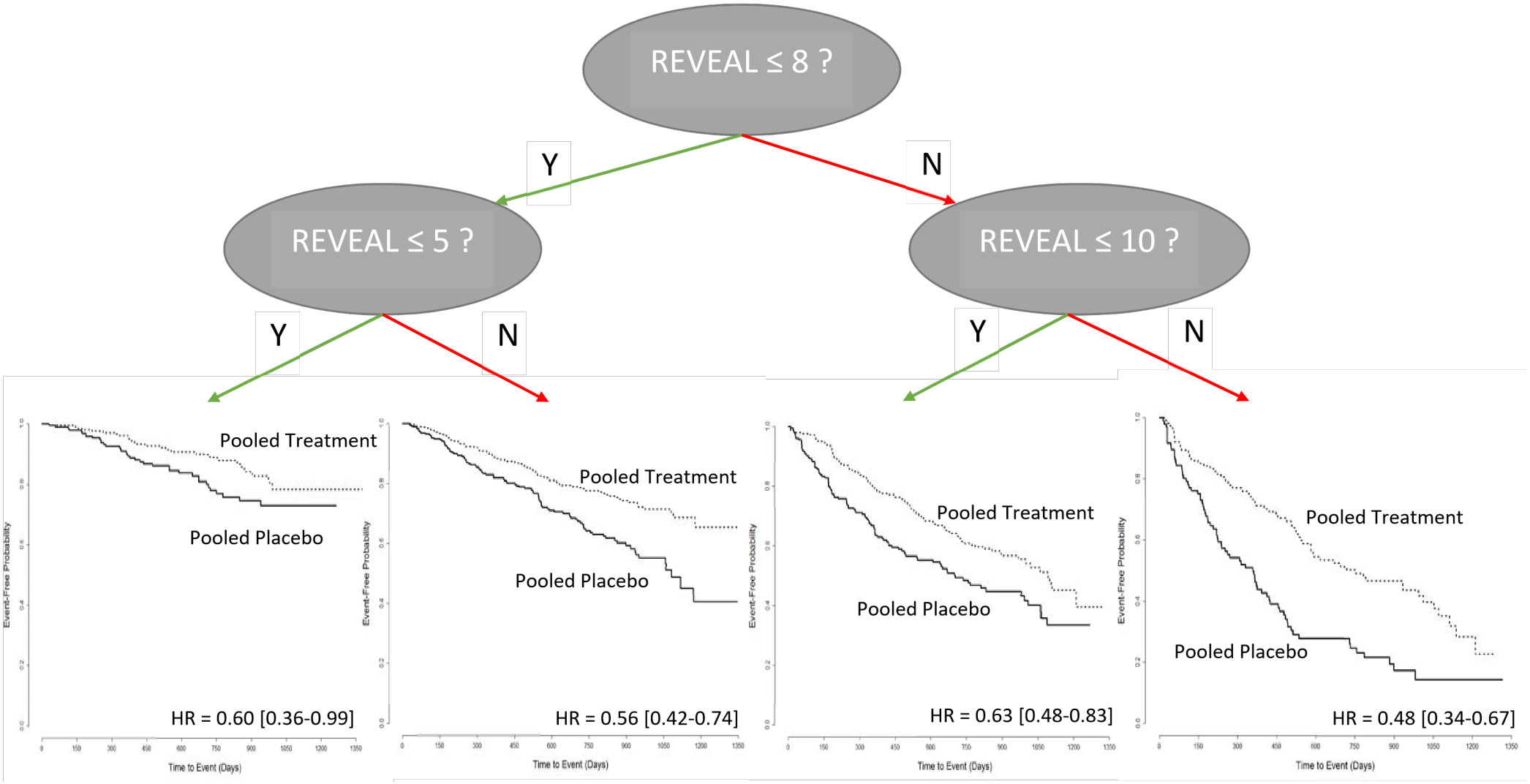


Figure 2.2: Survival Tree Analysis Applied with REVEAL 2.0 Risk Stratification.

When used in isolation, NYHA Functional Class identified only two statistically significantly different ranked groups ($p = 1.18 \times 10^{-9}$).

Hazard ratios between the pooled treatment and pooled placebo groups for reduction in clinical worsening rate were statistically significant for all risk groups identified, irrespective of risk algorithm used (all p-values < 0.05), demonstrating that even lower risk patients saw a treatment benefit and that bridging efficacy to these groups would be appropriate. Treatment effects were not significantly different between groups (i.e., there were no interactive effects between baseline risk and placebo versus treatment), but this retrospective study is not powered to determine interactive effects.

Impact on sample size and treatment time

Shown in Figure 2.3 are the results for sample size reduction (bar graph, y-axis left) and average treatment time reduction (superimposed line graph, y-axis right) for multiple enrichment strategies.

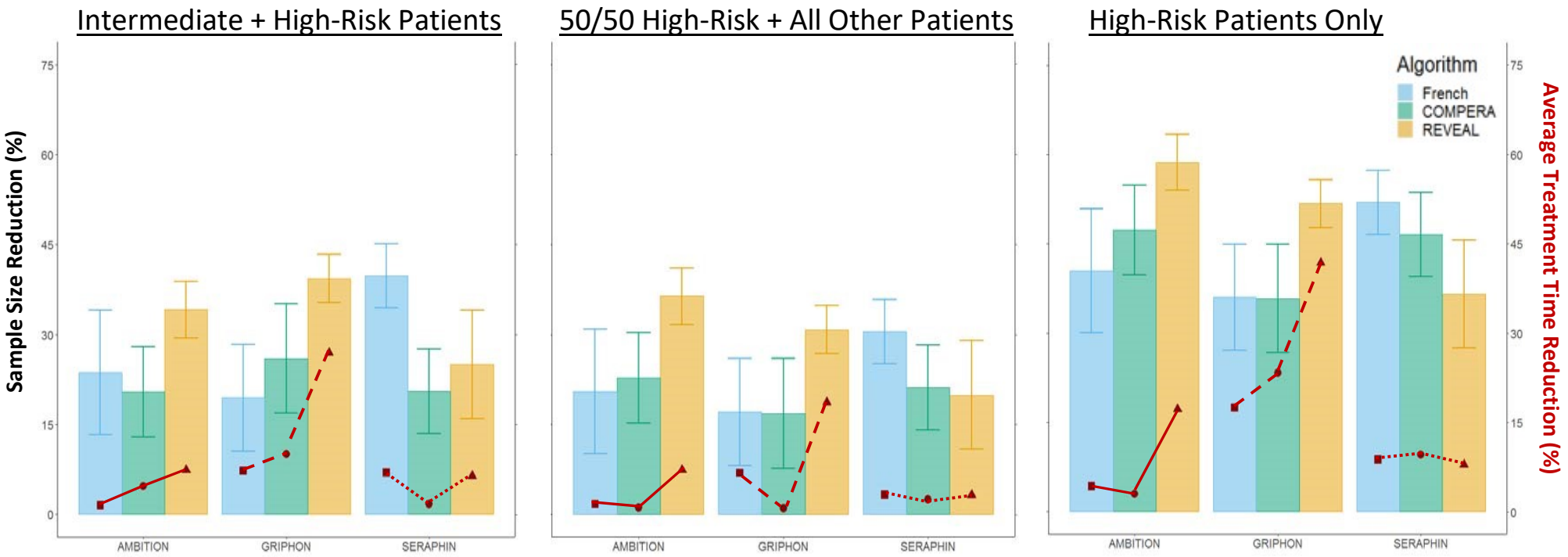


Figure 2.3: Estimated Sample Size and Treatment Time Reduction.

REVEAL 2.0 performed best for both reducing the total number of patients needed for enrollment and the average treatment time for all enrichment methods in AMBITION and GRIPHON. However, the French score, on average, outperformed both COMPERA and REVEAL 2.0 for all enrichment methods in SERAPHIN trial. This discrepancy is likely due to the use of a non-standard assay for measuring NT-proBNP in the SERAPHIN trial, which would compromise the accuracy of both COMPERA and REVEAL 2.0 for risk stratification in this trial, but not for the French score, as it does not use NT-proBNP to estimate risk. Table 2.6 provides the estimated screen-to-enroll ratios of each risk group per algorithm, as determined by the pooled dataset.

Enrichment Method	Risk Algorithm	Number Screened to Enroll 100 Patients
Intermediate and High Risk	COMPERA	170
	French	127
	REVEAL	263
50% High Risk/ 50% All Other	COMPERA	214
	French	259
	REVEAL	417
High Risk Only	COMPERA	427
	French	518
	REVEAL	833
None (Average of Original Trials)		124

Table 2.6: Screen to Enrollment Ratios per Enrichment Method.

REVEAL's screen-to-enroll ratio was highest (i.e., worst screening efficiency) for all enrichment methods. The French score using an intermediate-high enrichment strategy could achieve a comparable screen-to-enroll ratio to those of the original trials.

Impact on cost savings

Figure 2.4 shows the potential cost savings that GRIPHON may have benefited from if enrichment strategies had been used, as well as the ratio of screening cost to research (treatment) costs per enrichment strategy.

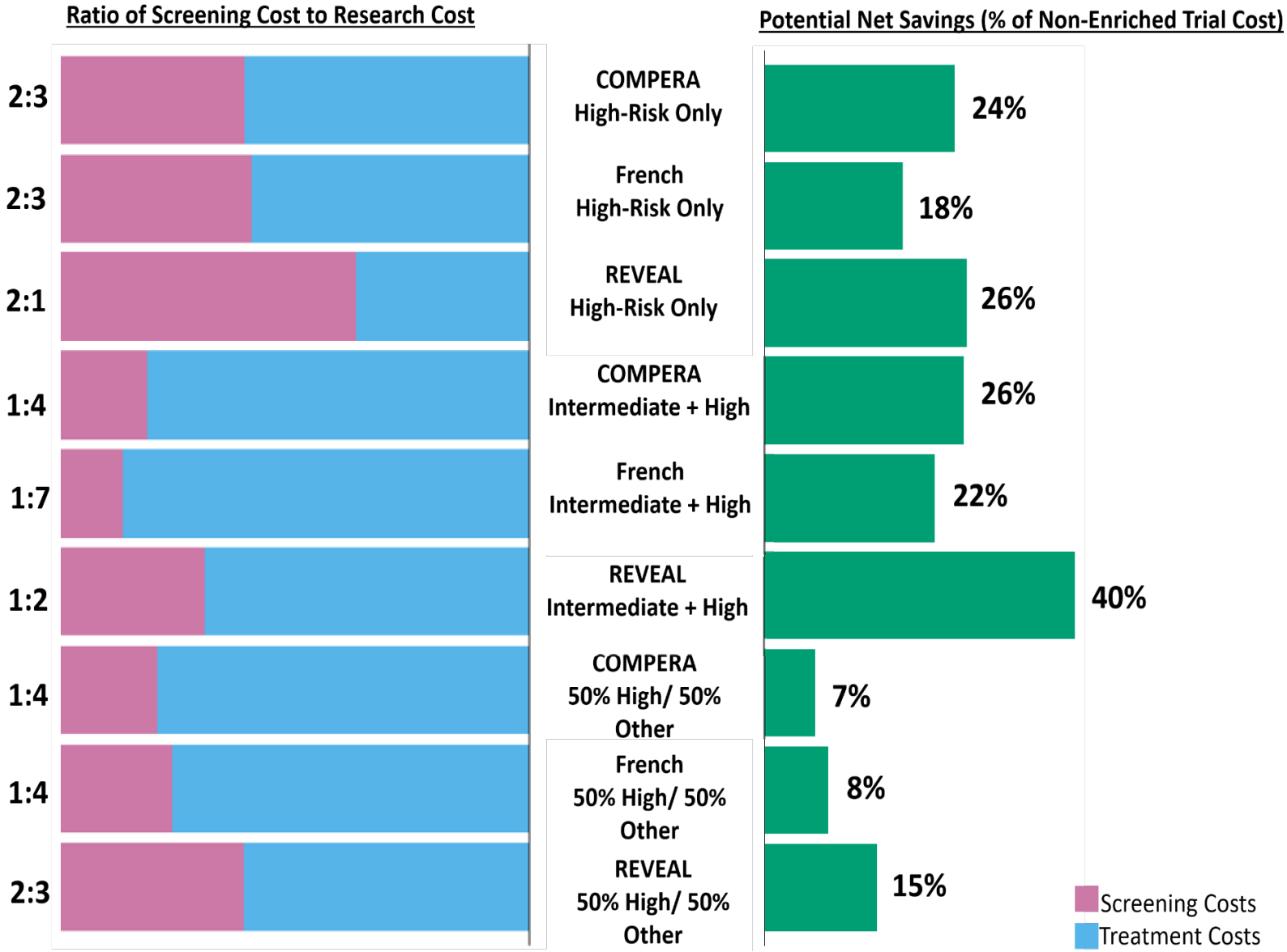


Figure 2.4: Cost Analysis for Applying Enrichment to GRIPHON.

Although all enrichment strategies reduced cost on average, the high-other strategy provided minimal net savings. Intermediate-high strategy provided the greatest financial benefit under REVEAL 2.0 (by reducing the total trial cost by 40%), and the high-risk only strategy reduced overall cost, but substantially increased screening costs.

Discussion

Clinical trial design in PAH has evolved into large, placebo-controlled studies focusing on a composite endpoint of clinical outcomes to determine therapeutic efficacy. However, such approaches are cumbersome and costly, and trial durations extend over many years in hopes of achieving the desired statistical power. The FDA supports using clinical trial enrichment, advocating the prospective use of patient characteristics to select a study population in which detection of a drug effect (benefit, or lack thereof) is more likely than in a broad patient population.(67) For any given desired power in an event-based study, the appropriate sample size depends on effect size and the event rate in the control group. Prognostic enrichment strategies are encouraged, not to affect the relative risk reduction but to increase the proportion of patients likely to experience a disease related endpoint, allowing for a higher number of events in a shorter time period, hence reducing overall sample size requirements.

In this analysis, all risk algorithms met the guidance criteria for prognostic enrichment: 1) they were shown to be prognostic of clinical worsening by ROC analysis, and 2) when applied using multiple enrichment strategies, they reduced the average estimated sample size compared with estimations made with the non-enriched population. Further, all risk groups were determined to have similar proportional pooled

treatment benefit after enrichment, supporting the FDA's statement that relative risk reduction is not affected with a prognostic enrichment strategy. As the FDA enrichment document cites, this concept has been demonstrated in multiple heart failure enrichment studies.(67)

As shown in Figure 2.3, REVEAL 2.0 outperformed COMPERA and the French score in two of the three clinical trials for all enrichment methods in reduction of relative sample size and trial length, but appeared less useful in SERAPHIN. One possible explanation for this variation is that REVEAL 2.0 and COMPERA were both optimized for a different NT-pro BNP analytical assay from the one used in the SERAPHIN trial.(52) Specifically, the SERAPHIN NT-pro BNP measurements were determined from an enzyme immunoassay rather than a chemiluminescence immunoassay, meaning its range and scale do not translate to the cut-points used in COMPERA and REVEAL 2.0. This leads to all SERAPHIN patients assessed appearing to have a "high-risk" NT-pro BNP level per these two algorithms. By comparison, the invasive French score does not consider NT-pro BNP for its algorithm and therefore could not be skewed by this value, which may be why it performed better than both COMPERA and REVEAL 2.0 at reducing sample size in simulations for SERAPHIN. This demonstrates the need for standardization of choice of biomarker tested and its range of values in risk assessment tools.

This result also leads into an important discussion about ease of use and data collection for risk-driven screening at trial baseline. REVEAL 2.0 requires 14 clinical variables (with a minimum of seven), while COMPERA requires five, and the French score requires four. Although the number of variables for each algorithm differs, the

overall cost of each algorithm as a screening tool was similar since right heart catheterization was the biggest contributing factor to cost. Further investigations are warranted to explore whether REVEAL Lite(98) or the French noninvasive score(60) can identify the intermediate-high and/or high-risk only group with considerably fewer variables and no hemodynamic variables. While the ideal enrichment strategy is a single, inexpensive biomarker, as shown in additional analyses, hemodynamic variables, six-minute walk distance, and NT-proBNP used in isolation were not as predictive as any of the three multivariable risk algorithms. Use of these singular biomarkers were unlikely to substantially reduce costs versus REVEAL 2.0, as confirmation of PAH diagnosis by right heart catheterization would still be required. Further, although results demonstrate that REVEAL 2.0 performed best in terms of cost-savings and precise prediction, if all variables considered in this analysis cannot be measured, it's recommended to use the risk algorithm that best fits the available data to avoid inaccurate prognosis of screened patients.

The cost-savings exercise conducted with GRIPHON data showed that intermediate-high risk enrollment as determined by a REVEAL score of ≥8 provides the greatest financial gain. This enrichment strategy had the triple benefit of (a) reducing the total number of enrolled patients, (b) reducing the average treatment time per enrolled patient, and (c) keeping the screening costs modest. For this analysis, most of the cost per enrolled patient stemmed from distributing study drugs over a long period of time. This estimation of other research costs was conservative, as it did not account, for example, for regularly scheduled lab tests, multiple right heart catheterizations, and the cost of personnel, all of which can vary considerably per trial and geographic

location.(99) For a comparison of scale, this economic analysis estimates an average cost of $17,000 USD per enrolled patient for 26 weeks of treatment, whereas current estimations for costs per enrolled patient in international trials with a median trial time of 26 weeks were $31,802 per patient as a low estimate.(100) This further illustrates the importance of keeping trial sizes small and treatment time short, even when requiring a more involved screening process. PAH clinical trials tend to be international, multi-site endeavors to reach necessary sample sizes. Economic research cited by the FDA established that international trials can benefit from lower cost of clinical procedures found abroad, although benefits in terms of the cost of study drugs is uncertain.(99) The estimates of the studied drug (Selexipag) used for analysis are based on previously published literature specific to its 2017 market price submitted by the manufacturer in the United States and Canada.(97) Specific Selexipag cost data in countries where costs of clinical trials are far lower, such as China and Russia, could not be found. While these estimates may not be entirely reflective of total clinical trial cost, especially for sites outside the United States, the analysis illustrates a proportional reduction in cost that would translate to international sites due to a reduction in treated patients and treatment time.

Currently, the FDA has no pharmacological concerns with bridging treatment efficacy established in a higher risk PAH group to treat lower risk PAH patients. The current understanding of PAH disease state and pathophysiology points to maintenance of a treatment effect regardless of a patient's individual risk of morbidity or mortality. However, it is unclear, though likely, that risk itself determines the level of treatment response. While prognostic enrichment is defined as selecting a patient population more

likely to experience a clinical event or endpoint, predictive enrichment seeks to specifically select patients who are more likely to be treatment responders. However, as stated in the FDA guidance document for trial enrichment, there is no absolute guarantee that prognostic enrichment and predictive enrichment are mutually exclusive. Further, when bridging treatment to lower risk patients, clinicians should consider risk-benefit tradeoffs, especially regarding how side effects may reduce quality of life, even when alleviating symptoms.

It's possible that risk-based prognostic enrichment also accomplishes the goal of predictive enrichment, by selecting for a patient population that experiences an effect that would not be present in an unselected population. Such a result was not supported by this analysis but cannot be ruled out for PAH drugs of differing mechanisms, as all investigated trials tested vasodilators. Further, an enriched patient cohort could experience a reduced rather than greater treatment response. This is possible specifically when a treatment cannot have a therapeutic effect quickly enough to slow the deterioration of a high-risk patient. Many cardiovascular drugs not expected to rapidly improve heart function still achieved approval by demonstrating efficacy in very ill patients with rapid deterioration and high mortality.(101) Therefore, it is not clear when it is too late to clinically intervene, especially with the limited pilot data that precedes a Phase III trial, and prognostic trial enrichment should still be considered.

The rarity of algorithm-identified risk groups is a serious challenge for enrichment. While a high-risk-only strategy provides the biggest reduction in sample size, the rarity of these patients makes screening prohibitive, not only financially but also in sheer number of available patients to screen. This analysis found that screen-to-

enrollment ratios were reflective of the real-world, as they matched the availability of an enriched population in the REVEAL registry, although rounding of the COMPERA score affected the apparent availability of its high-risk population.(61) These estimates are primarily meant to illustrate the relative difficulty of identifying a subpopulation within an already rare disease population (global prevalence of PAH are estimated at 15 cases per million). As stated above, PAH clinical trials typically require multiple sites across several countries to achieve adequate sample sizes and this is expected to still be necessary with an enrichment strategy. At this time, the need for additional sites to increase screening numbers is not expected with an intermediate-high risk strategy, but the possibility should be considered. Feasibility surveys should, therefore, be considered to determine the availability of intermediate and high-risk patients prior to site recruitment to mitigate costs.

This analysis assumed no pre-screening for estimation of screen-to-enrollment ratios. Pre-screening via electronic health records may allow a high-risk-only enrichment strategy to become more viable.(100) The feasibility of a high-risk-only enrichment strategy will become increasingly important as the sophistication and treatment costs of PAH (upwards of $10,000 USD per a 12 week period) clinical trials are increasing.(12) Trials with particularly high costs per enrolled patient, such as those studying dual or triple combination therapies, will substantially benefit from a high-risk-only enrichment strategy, specifically using REVEAL 2.0, which provided the greatest sample size and treatment time reductions with appropriate inputs.

Finally, the overall availability of intermediate- and high-risk patients to be enrolled per year also contributes to overall trial length. While PAH trials typically expect

to enroll roughly 200 to 350 patients per year, intermediate- and high-risk patients only compose an estimated 35-50% of the total current registry population.(61) Therefore, enrollment efforts are likely to slow, not entirely offset by the reduction in the total number of patients needed for enrollment. However, an enriched study will still stand to benefit from more primary endpoint events occurring at a quicker pace, as shown in this analysis with reduced treatment time. Therefore, trials may still benefit from a reduction in needed patient-treatment years. Future studies will investigate how to identify an enrichment strategy more precisely based on clinical trial simulations that account for availability of patients.

These findings address the limitations of current PAH clinical trials by demonstrating the benefit of risk stratification of patients with validated scales in PAH at baseline for optimizing enrollment. These data demonstrate for the first time the efficacy of established PAH risk prediction algorithms in selecting patients most likely to experience clinical events. When applied retrospectively to contemporary PAH clinical trials, this patient enrichment strategy reduced the enrolled/Intent-to-Treat population size required to detect a statistically significant treatment effect. Thus, results from this study establish that *a priori* risk stratification maximizes the likelihood of observing a statistically significant treatment effect with a smaller study population and should be considered in future study designs. Accordingly, the application of this approach to clinical study design is, in turn, expected to increase efficiency of successful PAH randomized clinical trials.

Limitations

This study had several limitations. Due to missing clinical variables at baseline, COMPERA and REVEAL 2.0 were not evaluated with all recommended variables. Further, all risk algorithms were optimized for predicting 1-year mortality rather than clinical worsening. COMPERA was applied without employing the rounding methods used with the algorithm applied to clinical risk stratification. Overall, not rounding has little effect on the overall functionality of the algorithm, as all inputs and calculations are the same. However, rounding could profoundly affect COMPERA's prognostic performance, as ranking accuracy almost always suffers when scales become less precise.

Relative algorithm accuracy was determined based on a pooled placebo group with differing definitions for clinical worsening. Given the nature of event adjudication, namely the lack of available source data surrounding worsening symptoms, it is infeasible to create a common clinical worsening endpoint to be used in all trials. This limitation motivates the understanding of how different risk algorithms perform for different definitions of clinical worsening, given all the proper inputs for calculations, which were not available for this study.

While sampling from a trial population with early withdrawal, power simulations did not force a specific attrition rate. Sample size reduction estimates were therefore controlled by comparing with a non-enriched population without attrition, rather than the original trial's study size. We expect the percentage change to be similar overall. Random drop-out, drug intolerability, or lack of satisfactory clinical progress (but not an explicit adjudicated event) can all contribute to an underestimation of drug effect, which

could substantially increase the required sample sizes. However, use of a high-risk enrichment strategy is less likely to be affected by attrition rates, as events occur earlier in the trial. It may, in fact, increase the likelihood of patients experiencing a clinically worsening endpoint before any other factor leading to an early withdrawal. In addition, by assuming that all patients with early withdrawal were event-free and thereby reducing the estimated event-rate in each risk group, estimations of sample size reduction may be conservative. The sensitivity analysis of the receiver-operating curves demonstrates that this assumption produces a modest bias in the performance of each algorithm. However, underestimating sample size reduction is desirable for providing appropriate recommendations versus providing inflated estimations that later prove to provide no significant reduction in sample size.

Differences in the statistical design of each study make it difficult to compare reductions in number of patients required across trials. Its promising that a significant reduction in relative sample size was shown for all trials, and overall the relative reduction in sample size was very similar across trials within a specific enrichment strategy (such as enrolling high risk only patients). However, these values may not translate directly for different statistical designs. It is difficult to speculate on the exact sample size savings for any given trial as this can depend on multiple factors, such as background therapy and other selection criterion.

In terms of bridging treatment efficacy, given the broad range of etiologies even for WHO Group I PAH, it is important to balance a trial's patient cohort appropriately such that it is representative of the intended drug indication. Our analysis demonstrated a treatment effect in all REVEAL risk populations, even though etiology contributed to

risk stratification, providing preliminary evidence that there are no concerns about bridging efficacy between risk groups of different etiological proportion. However, this concept warrants further investigation.

There are two major limitations to our economic analysis. First, increase in clinical trial costs due to the need for additional sites for screening was not considered. Next, all costs were based on estimates of clinical procedures and cost of the study drug (Selexipag) in the United States. While a proportional reduction in cost is still expected in other countries, it is speculative and likely not as large as estimated here.

Lastly, the studied population may not be representative of the risk in the entire trial population. Missing clinical variables for baseline risk assessment are assumed to be missing at random, but there may be some underlying cause.

Conclusion

Use of risk prediction algorithms as a prognostic enrichment tool must be validated in prospective clinical trials. This preliminary retrospective study demonstrates that such enrichment would reduce needed enrollment size and the duration of treatment and observation. This has many significant patient benefits, such as reducing the duration of treatment with placebo and improving time-to-market for potentially life-saving medications. Further, the financial burden of future PAH clinical trials can be reduced by improving trial efficiency, allowing drug developers to re-invest savings into research and drug innovation.

Chapter 3 : Risk-Benefit Tradeoff of Combination Therapy in Low-Risk PAH Patients

Introduction

As stated in Chapter 1, treatment strategies for pulmonary arterial hypertension are based on risk stratification at the time of initiation therapy and along the course of the disease.(10) The Ambrisentan and Tadalafil in Patients with Pulmonary Arterial Hypertension (AMBITION) clinical trial showed that initial combination therapy was associated with a 50% reduction in the risk of clinical failure vs pooled monotherapy in PAH patients.(42) This treatment effect was detected in both patients of low and moderate-to-high disease severity as defined by their New York Heart Association (NYHA) functional class. Prior to the results of this trial, it was unclear to clinicians whether low-risk PAH patients should be treated with upfront oral combination or monotherapy. This trial impacts a significant percentage of PAH patients given that about 15-30% of registry populations are deemed “low-risk” (<5% risk of 1-year mortality) at start of PAH therapy.(61)

As discussed in Chapter 2, risk scores are a viable tool for clinical trial enrichment, as they can precisely delineate low from intermediate or high-risk patients. Use of machine-learning methods shows promise in further improving the accuracy of risk stratifying PAH patients.(65) *Frost et al* initially published application of risk stratification to the AMBITION clinical trial data using the REVEAL [1.0] score, and concluded that

patients of all risk levels benefited from combination therapy, with no additional safety concerns across risk levels.(58) However, REVEAL [1.0] is now considered outdated, and more accurate risk stratification could reveal a significantly different risk-benefit tradeoff for low risk patients. Specifically, due to the increased risk of side effects caused by polypharmacy, as well as increased costs of combination therapy, ideally low-risk PAH patients would not need upfront combination therapy. As discussed in Chapter 2, risk scores are a viable tool for clinical trial enrichment, as they can precisely delineate low from intermediate or high-risk patients. Use of machine-learning methods shows promise in further improving the accuracy of risk stratifying PAH patients.(65) *Frost et al* initially published application of risk stratification to the AMBITION clinical trial data using REVEAL [1.0] score, and concluded that patients of all risk levels benefited from combination therapy, with no additional safety concerns across risk levels.(58) However, REVEAL [1.0] is now considered outdated and it is of significant interest to determine if improved risk stratification methods would draw the same conclusion.

This analysis assess the validity a published Bayesian network model (Pulmonary Hypertension Outcomes Risk Assessment or PHORA) within the AMBITION clinical trial.(65) This study assesses both the benefit of upfront mono- versus combination therapy as well as potential risk tradeoff in low-risk patients compared to those of higher risk. It was hypothesized that although some benefit with upfront combination therapy over monotherapy may exist for low-risk patients, increased risk of side effects (i.e. adverse events or AEs directly attributable to the experimental drug) on combination therapy versus monotherapy makes this treatment strategy less tenable.

Methods

The AMBITION clinical trial dataset was used for this analysis, the details of which have been previously published.(42) Briefly, this was a randomized, double-blind event-driven study whereby treatment naïve PAH patients were assigned to either combination therapy (Ambrisentan plus Tadalafil) or Ambrisentan monotherapy or Tadalafil monotherapy at a ratio of 2:1:1, respectively. The primary endpoint was time to first adjudicated clinical failure event, defined as: death, hospitalization due to worsening PAH (including initiation of parenteral prostanoid therapy, lung or heart and lung transplantation, or atrial septostomy), disease progression (a decrease of more than 15% from baseline in 6-minute walk distance (6MWD) or 6MWD combined with WHO FC III or IV symptoms) or unsatisfactory long-term clinical response (any decrease from baseline in 6MWD at two consecutive visits separated by 2 weeks and WHO FC III symptoms at two consecutive visits separated by 6 months). Lastly, adverse drug events were analyzed for drug safety.

The original AMBITION clinical trial data was analyzed through direct collaboration with the U.S. Food and Drug Administration (FDA). Patients analyzed (N = 500) included only those specified in the modified intention-to-treat population per the original trial publication, excluding patients with evidence of left heart disease that did not meet the amended entry criteria.(42) As a previous retrospective analysis used REVEAL [1.0] to detect a low-risk population, the current study aimed to demonstrate the improved stratification of PHORA to reclassify patients into low-risk versus intermediate-high risk, as well as determine if differences in treatment response exist between risk groups. A description of the design of these two scores, including inclusion

and exclusion criteria, has been published previously.(62; 65) Briefly, REVEAL [1.0] is a risk-score based on a Cox proportional hazards model of one-year survival, while PHORA is a tree-augmented naïve Bayesian network that estimates a probability of survival in one-year. Of note, PHORA employs the same variables and cut-off values as the contemporary REVEAL 2.0 risk score.

The AMBITION dataset used the following variables to calculate risk in both REVEAL [1.0] and PHORA: age, gender, etiology of PAH, (resting) heart rate, systolic blood pressure, NYHA functional class, 6MWD, estimated glomerular filtration rate (eGFR), NT-proBNP, mean right atrial pressure (mRAP), and pulmonary vascular resistance (PVR). Missing variables at baseline included diffusing lung capacity and hospitalization six months prior to randomization. Presence of pericardial effusion was derived from patients' medical histories at baseline, imputed as "Yes" if present in medical history, but left as missing if not mentioned. Estimated glomerular filtration rate at both baseline and 16-week reassessment were calculated from serum creatinine levels per the *Levey et al* equation.(91)

For comparison to the *Frost et al* article, which first explored risk stratification for combination therapy in PAH, risk groups were formed as defined by that publication for the REVEAL score (< 6 low-risk, ≥ 6 intermediate-high risk).(58) PHORA risk groups were formed per the ERS/ECS guidelines (<5% predicted 1-year mortality was low-risk, ≥ 5-10% intermediate-high risk).(10)

Receiver-operator curves (ROC) and area under the curve analysis is used to compare the performance of the two risk scores (REVEAL [1.0] and PHORA). To ensure treatment is not a confounding factor of mortality and baseline risk, only patients

in the placebo arm are used for ROC analysis. Net reclassification index (NRI), defined by Equation 9, was calculated to determine if PHORA had an improved ability to discriminate between low-risk and intermediate-high risk patients, again only using placebo arm patients.(102)

$$NRI = P\left(N_{high\ to\ low}\middle|event\right) - P\left(N_{low\ to\ high}\middle|event\right) + P\left(N_{high\ to\ low}\middle|nonevent\right) + P\left(N_{low\ to\ high}\middle|\ nonevent\right), \quad (9)$$

where *P* is the probability or percentage, $N_{high\ to\ low}$ indicates the number of patients who are reclassified from intermediate-high risk to low risk, $N_{low\ to\ high}$ indicates the number of patients who are reclassified from low risk to intermediate-high risk, *event* indicates death within one-year, *nonevent* indicates no death within one year.

In the AMBITION clinical trial, the primary endpoint, time to clinical failure, differed from other PAH event-driven trials (e.g. GRIPHON and SERAPHIN). AMBITION's primary endpoint included an event for "unsatisfactory clinical response", which is an event with limited and unclear clinical relevance, as it specifies that any decrease in six-minute walk distance and persistence of NYHA class III symptoms after 6 months constituted a treatment failure. Therefore, for this study's primary analysis, time to clinical worsening (which includes only death, hospitalizations for worsening PAH, and disease progression) was analyzed for each risk group (censored at one-year and at final assessment visit).(16) Mortality (all-cause death) censored at one-year and follow-up was a secondary analysis. Placebo patients who switched to combination therapy prior to the primary endpoint were censored at the day of start of combination therapy. Kaplan-Meier curves were generated for time to clinical worsening per risk group and treatment arm as well as time to death. Cox proportional hazards were used

to determine differences between combination and monotherapy within each risk group, using a 5% alpha (α) level to define statistical significance. Additionally, time to clinical failure (i.e. time to clinical worsening plus unsatisfactory clinical response) was analyzed per risk group at one-year and follow-up as a supplementary analysis.

To examine the risk of upfront combination therapy, the proportion of PHORA low-risk patients experiencing significant treatment-emergent adverse events (AEs) were analyzed. Treatment-emergent significant AEs were defined as those that lead to the withdrawal of the investigational product (i.e. combination therapy or monotherapy) and were determined by the investigator to be related to the investigational product. This also included events that had the potential for liver injury, defined by an elevation of liver aminotransferases of 5 x upper limit of normal (ULN).(42) AEs as determined by these criteria were reviewed by PAH clinical experts to determine their clinical relevance and broadly categorize event types. The proportion of patients experiencing these events in each treatment arm (Ambrisentan monotherapy, Tadalafil monotherapy, and combination therapy) are reported. Barnard's (one-sided) exact test was used to determine if a significantly smaller proportion of low-risk patients on monotherapy experienced adverse events versus patients on combination therapy.

Finally, risk-benefit tradeoff was examined using Cox proportional hazards to determine differences in time to either first AE or first clinical worsening event between combination and monotherapy within each risk group, using a 5% alpha (α) level to define statistical significance.

Results

Table 3.1 provides a summary of total patients classified as low and intermediate-high risk by both REVEAL [1.0] and PHORA at baseline and 16-week reassessment. Both PHORA and REVEAL 1.0 were prognostic for one-year (PHORA AUC = 0.73 and REVEAL 1.0 AUC = 0.73; Figure 3.1). Net reclassification tables for one-year mortality events are shown in Table 3.2.

Algorithm	Time Period	Low Risk (N, %)	Intermediate-High Risk (N, %)
REVEAL 1.0	Baseline	75 (15.0%)	425 (85.0%)
PHORA	Baseline	216 (43.2%)	284 (56.8%)

Table 3.1: Risk Populations at Baseline.

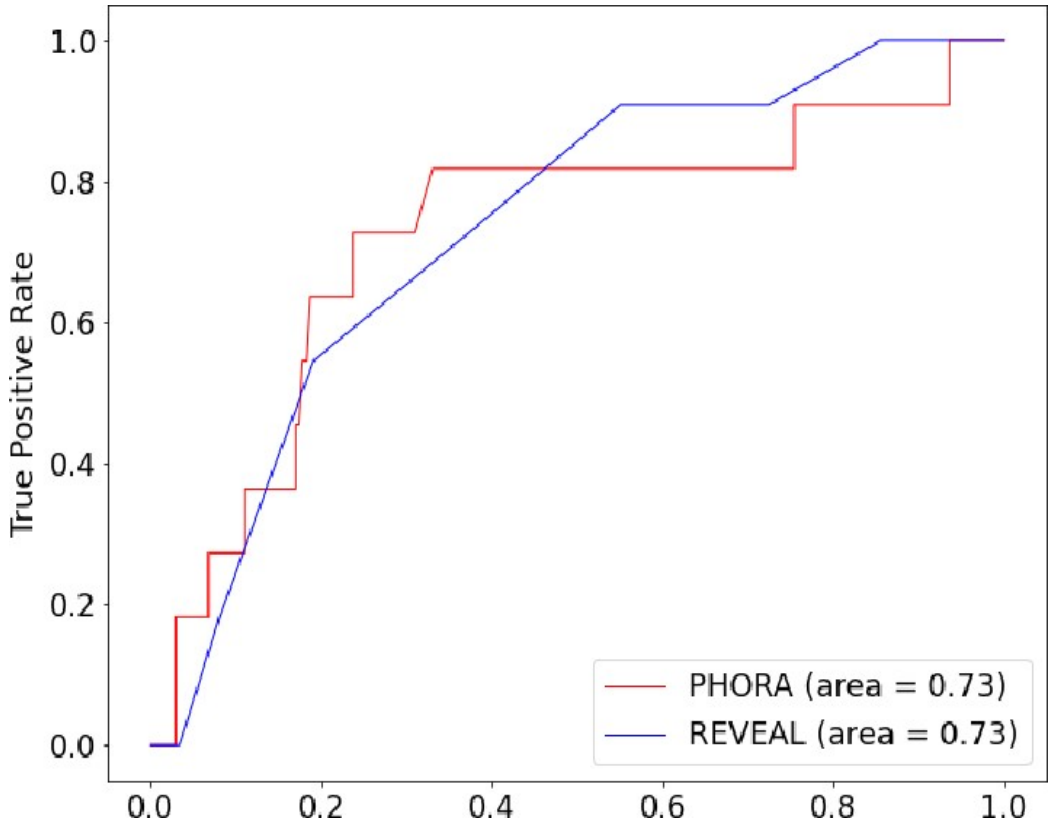


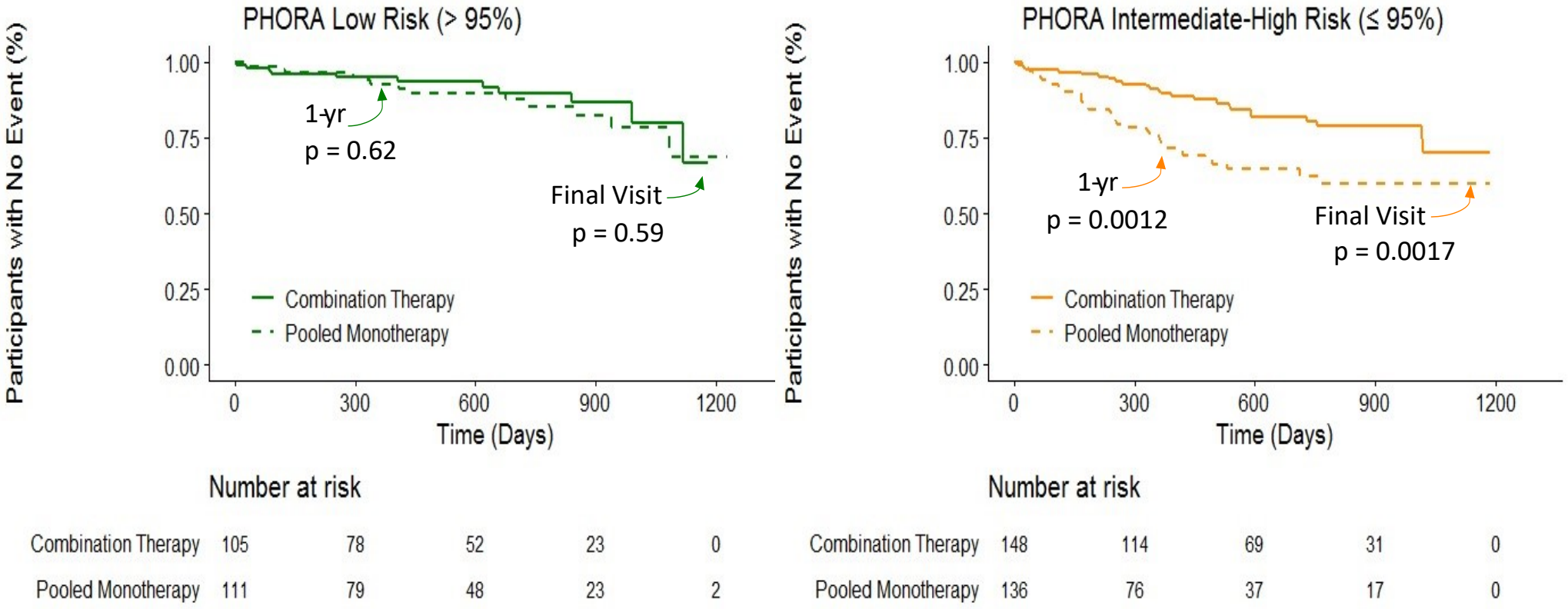
Figure 3.1: Receiver-operator curve for REVEAL [1.0] versus PHORA at Baseline.

			REVEAL 1.0	
			Intermediate-High	**Low-Risk**
PHORA	**Intermediate-High**	**With Death**	9	0
		W/o Death	127	0
	Low-Risk	**With Death**	2	0
		W/o Death	34	75
Total			172	75
Reclassification For Events			-0.0116 (-2/172)	
Reclassification for Non-events				0.453 (34/75)
Net Reclassification Index			0.441	

Table 3.2: Net Reclassification Analysis Between PHORA 1.0 and REVEAL 1.0

Although the ROC-AUC were nearly equivalent between the models, the overall net reclassification index for PHORA was positive (NRI = 0.441), indicating improved discrimination between low versus intermediate-high risk compared to REVEAL 1.0. See discussion for further comment.

Kaplan-Meier curves for time to clinical worsening per risk group and treatment arm is shown in Figure 3.2.



Number at risk

Combination Therapy	105	78	52	23	0
Pooled Monotherapy	111	79	48	23	2

Number at risk

Combination Therapy	148	114	69	31	0
Pooled Monotherapy	136	76	37	17	0

Figure 3.2: Combination versus Monotherapy for Time to Clinical Worsening in Low-Risk and Intermediate-High Risk Patients.

PHORA's low risk group did not have a significant delay in clinical worsening when comparing upfront combination therapy to monotherapy, censored at one-year (HR = 0.75, p = 0.62), nor at final assessment (HR = 0.80, p = 0.59). This contrasts with the results of *Frost et al*, by which REVEAL's low risk group was concluded to have a significant benefit on combination therapy. However, PHORA's (>5% predicted risk of mortality) intermediate-high risk group saw significant treatment benefits when censored at one-year (HR = 0.34, p = 0.0012) and final assessment visit (HR = 0.45, p = 0.0017). Using the original trial primary endpoint, time to clinical failure, PHORA's low-risk combination therapy group still did not experience a significant difference (HR = 0.67, p = 0.22; see Appendix). No risk groups (as determined by PHORA) saw a significant treatment benefit (all p-values > 0.20) for mortality when censored at one-year nor at follow-up (Figure 3.3). Table 3.3 summarizes the first AEs leading to product withdrawal experienced in the low-risk groups for PHORA.

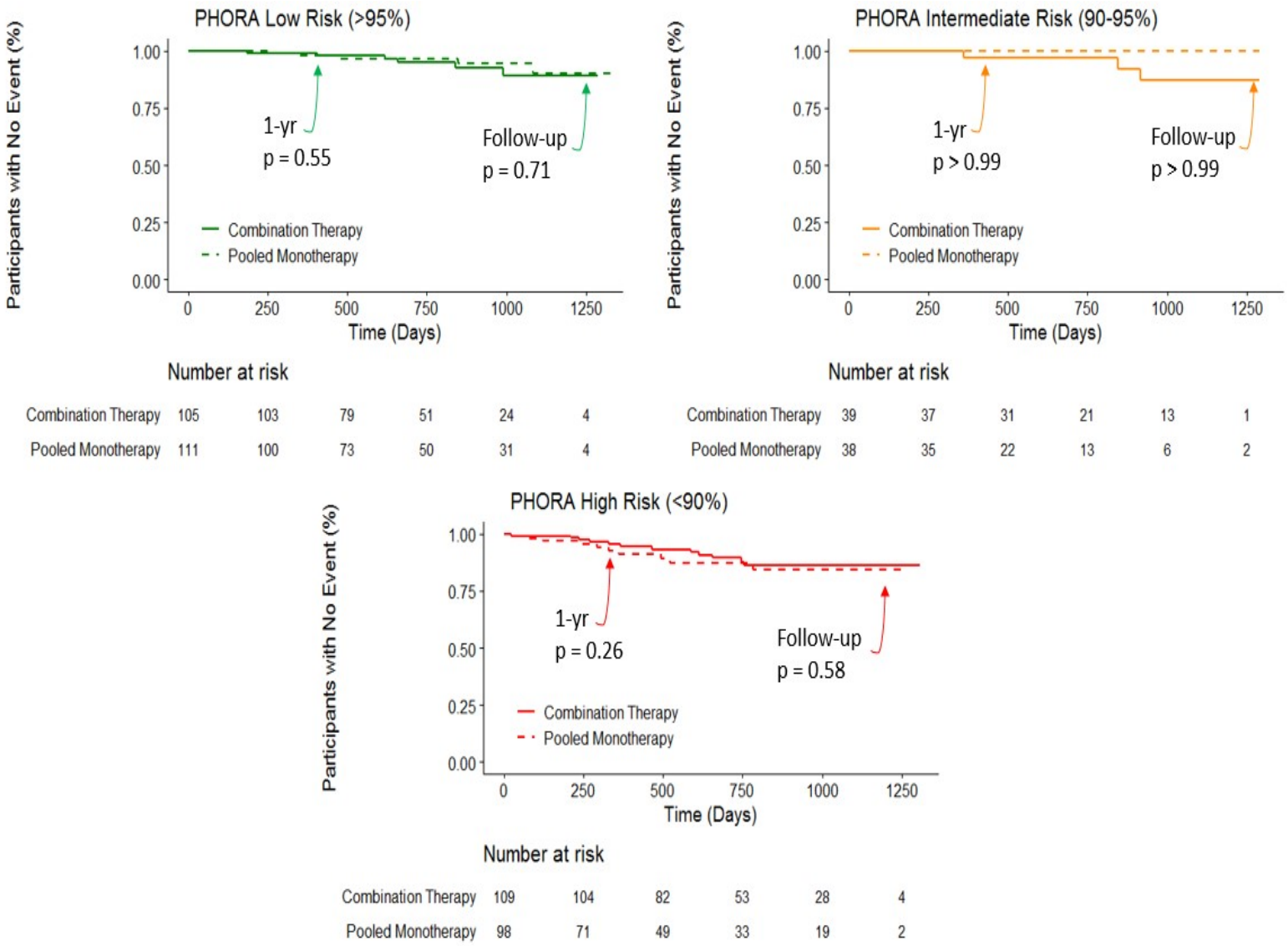


Figure 3.3: Mortality Events for Combination Therapy vs Monotherapy, Separated By Low, Intermediate, and High Risk

Adverse Event Type	Ambrisentan Monotherapy (N = 58)	Tadalafil Monotherapy (N = 53)	Combination Therapy (N = 105)
Pulmonary Edema	1.7% (1)	0.0% (0)	0.0% (0)
Potential for Liver Injury	0.0% (0)	1.9% (1)	1.0% (3)
Hypersensitivity	0.0% (0)	0.0% (0)	1.9% (2)
Peripheral Edema	3.4% (2)	1.9% (1)	2.86% (3)
Myopathy/Myalgia	0.0% (0)	1.9% (1)	1.0% (1)
Renal Failure	1.7% (1)	0.0% (0)	0.0% (0)
Hypoxia	0.0% (0)	0.0% (0)	1.0% (1)
Nasal Congestion	0.0% (0)	0.0% (0)	1.0% (1)
Multiple AEs (general intolerance)	0.0% (0)	1.9% (1)	3.8% (4)
Total Percentage Experiencing Adverse Events Leading to IP Withdrawal	6.9% (4)	7.5% (4)	14.2% (15)
Total Percentage Experiencing Adverse Events Leading to IP Withdrawal (Monotherapy vs Combination Therapy)	7.2% (8)		14.2% (15)
p-value (Barnard's Exact Test, one-sided)	0.056		

Table 3.3: PHORA Low-Risk Patients First Adverse Event Leading to Treatment Withdrawal.

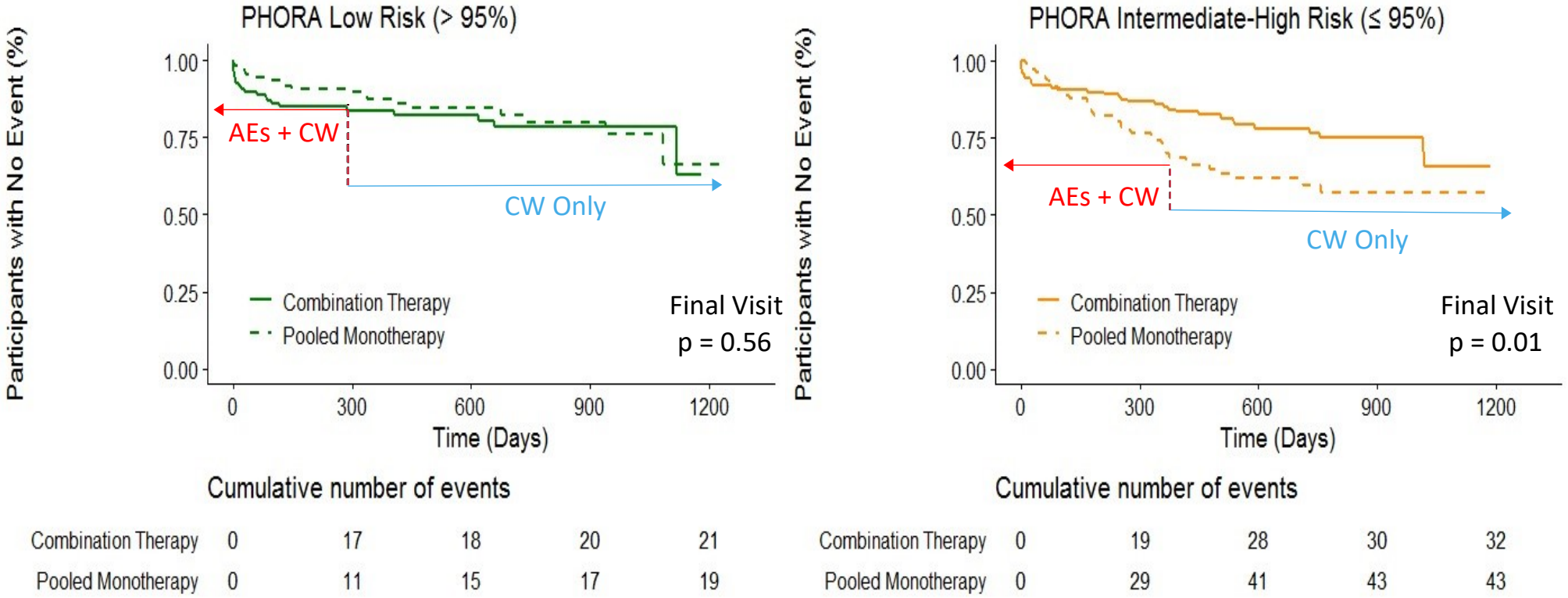
The proportion of patients experiencing an adverse event in the combination therapy group was twice as high compared to that of monotherapy (14.2% versus 7.2%, respectively; p = 0.056). Table 3.4 summarizes the first clinical worsening events experienced in the low-risk groups for PHORA low-risk.

Clinical Worsening Event Type	**Ambrisentan Monotherapy (N = 58)**	**Tadalafil Monotherapy (N = 53)**	**Combination Therapy (N = 105)**
Death (all-cause)	2	0	5
Hospitalizations due to worsening PAH	5	1	2
Disease Progression	4	2	4
Total Events	11	3	11

Table 3.4. PHORA Low-Risk Patients First Clinical Worsening Events.

Low-risk patients on combination therapy experienced death more often as their first clinical worsening event, while patients on monotherapy experienced hospitalization or disease progression more often, but these differences were not statistically significant.

When adverse events and clinical worsening were combined in time-to-event analyses, PHORA low-risk patients did not see a significant difference between combination therapy and monotherapy when censored at final assessment visit (HR = 1.20, p = 0.56). PHORA intermediate-high risk patients, in contrast, still saw a significant treatment benefit even in the presence of AEs at final assessment visit (HR = 0.56, p = 0.010). Kaplan-Meier curves for this time to event analysis are shown in Figure 3.4.



Cumulative number of events

Combination Therapy	0	17	18	20	21
Pooled Monotherapy	0	11	15	17	19

Cumulative number of events

Combination Therapy	0	19	28	30	32
Pooled Monotherapy	0	29	41	43	43

Figure 3.4: Combination versus Monotherapy for Time to Clinical Worsening or Adverse Drug Event in Low-Risk and Intermediate-High Risk Patients

All AEs leading to product withdrawal in the low-risk groups occurred on or before 288 days of treatment (median = 21 days for PHORA low-risk) and were largely categorized as general drug intolerance and/or hypersensitivity. For intermediate-high risk, all AEs occurred on or before 372 days (median = 28 days for PHORA intermediate-high risk).

Discussion

Contemporary PAH risk stratification tools allow for a more accurate study of the benefit of upfront combination therapy on low-risk patients. PHORA demonstrated a net positive reclassification index versus REVEAL [1.0]. This demonstrates the improved discriminatory power of Bayesian networks versus traditional risk calculators and allowed us to analyze treatment heterogeneity more accurately among treatment-naïve patients with pulmonary arterial hypertension. In this analysis, the number of patients identified as low-risk with PHORA is greater than the number identified with REVEAL 1.0. While REVEAL cut-offs use a score of < 6 and > 6 for low and intermediate-high, PHORA considers patients in risk categories per the ERS/ ESC guidelines. Of note, REVEAL 1.0 is a simple additive equation, whereby any patient with a single "higher risk" variable and no protective variables will automatically become intermediate risk. PHORA, on the other hand, weights variables by their interactions by virtue of its methodology. This allowed PHORA to identify significant more truly low-risk patients.

The results of the original AMBITION trial and retrospective analysis with REVEAL 1.0, supported the notion that all patients benefitted from upfront combination therapy.(58) However, improvements in PAH risk stratification using the PHORA

classification system demonstrates that this benefit is not necessarily clear for PAH patients characterized as low-risk. As stratified by PHORA, low-risk PAH patients did not have a significant delay in clinical worsening nor clinical failure with combination therapy. These results would not necessarily suggest that clinicians should no longer treat low-risk patients with upfront combination therapy, as PAH remains a progressive and deadly disorder even for this group. Since the mechanism of drug efficacy is maintained in the low-risk population (i.e. targeting multiple disrupted pathways that contribute to pulmonary vasoconstriction and cellular proliferation), given a sufficiently statistically powered study, low-risk treatment naïve patients with pulmonary arterial hypertension would likely significantly benefit from combination therapy. However, it's clear from the risk-benefit analysis that these reductions in clinical worsening do not outweigh the increased risk of intolerable side effects. Further, the substantial increase in financial cost of oral combination therapy makes this strategy less attractive for low risk patients. In financial analyses of healthcare burden on PAH patients, reduction of inpatient hospitalizations is key to offset the increased costs of pharmaceuticals. However, as low-risk patients are far less likely to experience hospitalizations in the first place, it is unlikely that the additional cost of upfront combination therapy versus monotherapy (e.g. $8,750 for 6 weeks of treatment, versus $6,050 for ERA monotherapy and $2,700 for PDE5 inhibitor monotherapy) outweighs the minimal gains in clinical worsening.(89)

While Chapter 2 asserted that establishing treatment efficacy initially in a higher risk population should not interfere with prescribing medications to lower risk groups, Chapter 2 also maintained that increased risk of side effects may outweigh benefits of

combination treatment for certain groups. Clinicians should consider the multiple facets of their patients' lives when choosing a treatment plan, including potential for adverse side effects and how these side effects reduce quality of life. While most adverse effects observed in this study were not as harmful as PAH disease progression, if the side effect led the investigator to the discontinue the medication, it is likely due to the wishes of the patient. Therefore, the AEs identified in this study, which are adjudicated by the investigator to be directly caused by the drug and lead to discontinuation, are the most important events to consider right after the clinical worsening events themselves. Specifically, no significant benefit nor harm was found for the low-risk PAH patients, but the point-estimate of the hazard ratio indicated a greater propensity for harm (i.e. HR = 1.2). This is contrasted with results for intermediate-high risk patients where the hazard ratio was statistically significant and demonstrated reduction of overall harm, from either clinical worsening or adverse events (HR = 0.56).

Side effects from combination therapy should be mitigated as much as possible to improve treatment compliance. According to a public meeting held by the U.S. Food and Drug Administration in 2013, in which approximately 85 patients with PAH or their representatives attended, many patients commented that they frequently found they had to discontinue a specific medication due to side effects or frequently tested "one drug at a time" in order to identify which therapies gave the greatest alleviation of PAH symptoms with the fewest side effects. The FDA stated in their concluding remarks of the meeting that: "Drug treatments that slow the progression of the disease are available for PAH patients; however, efficacy varies from patient to patient, and

significant side effects can limit benefits or preclude use of these medications." Therefore, it's clear that significant side effects impact patients' medical compliance.

This analysis found that the first year of combination therapy treatment is an important window for monitoring for adverse drug effects. Proportion of patients experiencing AEs in the low-risk combination therapy group were still relatively rare compared to monotherapy but occurred almost twice as frequently (14% versus 7%, $p = 0.056$). Appropriate alternatives to upfront combination therapy, including sequential add-on therapy or slow titration with regular monitoring could be a potential viable strategy for mitigating adverse side effects. In terms of choosing a first initial monotherapy, clinicians should weigh the risks of the drug with comorbidities (such as liver dysfunction or altered drug metabolism), hinderance on quality of life, and financial affordability in countries without universal health care.

In determining treatment efficacy, especially between different risk groups, the clinical trial endpoint should be relevant to all patients and well-defined. From a patient perspective, it is not meaningful for a drug to demonstrate efficacy on a primary endpoint that they are unlikely to experience in the first place. Further, composite endpoints essentially weigh all events, including mortality, equally, and prioritize the first event, obscuring which elements of the composite a treatment truly effects. For low-risk patients, this typically means that events like disease progression or failure to improve are over-emphasized, albeit infrequent. The infrequency of such events for low-risk patients makes the clinical worsening endpoint less relevant to them. Ideally, an endpoint for low-risk PAH patients would focus more on alleviation of symptoms, especially breathlessness, clinical maintenance of low-risk status, and no significant

increases in side effects that preclude treatment. Prognostic enrichment, therefore, be used in event-driven PAH clinical trials to determine treatment efficacy in an intermediate-high risk group and then later establishing efficacy in a low-risk population through a different set of endpoints.

Finally, this study does validate the crucial need for upfront combination therapy for intermediate and high-risk patients. As shown in the risk-benefit analysis, even with an increased risk of side effects, there is a clear clinical benefit for this higher risk group in reducing clinical worsening events. Interestingly, high risk patients did not confer a significant benefit in the delay or reduction of mortality events. This suggests that upfront oral combination therapy may not be sufficient for increasing lifespan in high-risk PAH patients, and that triple oral combination therapy or intravenous epoprostenol treatment is warranted. This agrees with the 2015 ERS treatment guidelines and further underscores the crucial need for accurate risk stratification to protect against the undertreatment or overtreatment of PAH patients.

Limitations

There are several limitations within this work. First, in the absence of a standard definition of a clinical worsening, elements of composite endpoints that have been used in other PAH clinical trials were chosen to allow for more comparable results, but the adjudication conducted in other PAH trials could not be exactly reproduced. Specifically, in the original trial, PAH clinical experts reviewed suspected clinical worsening events and determined if their symptoms warranted the definition of NYHA Class III to meet the criteria for clinical failure, although the patient may never have fully achieved this

classification such that they were recorded as such during follow-up visits. Second, not all patients had all variables required for risk stratification and the effects of missing data on algorithm accuracy have not yet been directly studied. All patients did, however, have the requisite seven of thirteen variables recommended for use in PHORA. Third, categorizing and combining AEs is ambiguous under the current MedDRA system and our investigation does not reflect an exhaustive analysis of the types of events. For this reason, no statistical analyses were performed on specific event types and only investigated the collective impact on time to clinical worsening or drug discontinuation.

Clinical practice would strongly benefit if a cut-off for a “very low risk” PAH group could be determined and used to study further differences in treatment response, but low sample size made this infeasible for this study. As evidence by the Net Reclassification, this more sophisticated algorithm could identify more ‘truly low risk’ patients that won't have a mortality event at the end of the year. Although PHORA is not yet the perfect risk stratification algorithm, this study demonstrates that improvements in risk stratification technology are crucial to improving understanding of PAH clinical treatment strategy. Advances in risk stratification, the focus of the next chapter, will allow for more focused analysis of low-risk patients in the future.

This is a retrospective analysis that is not powered sufficiently to claim a complete lack of treatment benefit in the low-risk group. Further investigations with real-world data are warranted to achieve an adequate number of events and statistical power to reduce chances of a false negative treatment benefit.

Conclusions

In summary, the overall safety of a drug, healthcare costs, potential for a treatment benefit, and patient reported outcomes must all be considered when choosing a treatment strategy. These results demonstrate that the overall benefit-risk tradeoff for low-risk PAH patients does not favor upfront oral combination therapy. This finding provides an opportunity for clinical practice to be re-evaluated, limiting treatment costs and drug side effects for 15-30% of the PAH patient population. Further, the need for more aggressive treatment for high-risk PAH patients in the form of triple combination or intravenous prostanoid therapy is strongly supported by this study.

Overall, this study demonstrates the clear importance of accurate risk stratification for PAH treatment strategy and the use of more sophisticated risk tools would allow providers to more accurately communicate the potential benefits and risks associated with PAH medications and tailor individual therapies. A larger sample of low-risk patients should be leveraged to further investigate concerns regarding the risk-benefit tradeoff of upfront combination therapy versus sequential add-on therapy or strategic dosing in low-risk WHO Group I PAH patients.

Chapter 4 : Beyond the Risk Calculator - Improving PAH prognosis with machine learning

Introduction

As stated in Chapter 1, risk calculators are considered the best clinical strategy at this time for guiding pulmonary arterial hypertension treatment and avoiding sudden clinical worsening. These risk calculators aim to sort patients from low to high risk based on clinical markers.(66) Current commonly used tools include three risk calculators developed based on registry data: COMPERA(59), French (FPHR)(60), and REVEAL 2.0(61). Each calculator uses a combination of hemodynamics, demographics, and/or laboratory values to assign a risk level to a patient based on predicted mortality: low risk, with a less than 5% one-year mortality rate, intermediate risk, with a 5-10% one-year mortality rate, and high risk, with a greater than 10% one-year mortality rate. However, there is currently no single, agreed-upon tool to determine risk and all common clinical risk tools based on registry data are at best fair-to-good (c-statistic 0.65-0.75) in their performance.(10; 61)

Further, most risk calculators specifically focus on metrics of right ventricular function or severity of symptoms but ignore other systemic biomarkers of right heart function. Right heart failure causes several negative downstream effects on the kidneys, liver, gut, immune system, many of which have been linked to worsening clinical outcomes and mortality.(103) Clinical guidelines recommend the use of liver and kidney metabolic panels for PAH patient assessment, but these laboratory values have yet to be

incorporated into a risk calculator.(10) Further, conditions such as hyponatremia and hyperbilirubinemia are recommended criteria for placement on the lung transplantation waiting list, yet these serum values are also not considered in PAH risk calculators.(104) Biomarkers from these organ systems may enhance prognostic power of clinical risk tools by highlighting patients with poor right heart adaptation specific to their individual organ needs rather than using “one-size-fits-all” ideal values for RV hemodynamics.

In 2019, the first Tree-Augmented Naïve (TAN) Bayesian network, based on the REVEAL 2.0 calculator, was validated in multiple PAH registry datasets (shown in Chapter 1).(65) Its performance for prediction of one-year mortality improved incrementally upon REVEAL 2.0, but it did not consider how an expanded feature set could improve prognostic performance.

In addition to improvements in performance, there are many other advantages to the use of Bayesian network modeling for clinical prognosis, which were reviewed in Chapter 1. Briefly, Bayesian networks allow the user to map clinical variables to a probability of an event, rather than giving a relative, ordinal risk level, which is less precise. Bayesian networks have several significant advantages over parameterized statistical methods, which must force certain assumptions about how clinical variables translate to risk, such as proportional hazards.(69) Lastly, the usability and accuracy of classic statistical tools become less reliable when specific measurements are missing from patients and cannot be reasonably imputed.(73; 105) This is a common problem with clinical data where a clinician’s time and resources can be very limited, or the cost of repeated invasive testing is prohibitive. Whereas a traditional risk score does not have a clear means of handling missing data, Augmented Bayesian Networks can substitute

the most likely variable given all others observed and expectation-maximization can be used when training the network with non-complete case data.(106) Lastly, Bayesian Networks can be represented with a graphical structure. This increases user confidence as the model is not a “black box” and interactions between variables, the outcome, and the strength of the relationship can be visualized.(107)

The goal of this study was therefore to consider an expanded set of clinical variables and develop a new Bayesian network from “the ground-up” rather than basing it entirely on another risk calculator. The goal of developing this new network was to improve upon performance of published risk calculators while reducing the number of needed variables, integrate use of common laboratory values into risk assessment, and improve upon the understanding of the interplay between clinical variables and short-term mortality outcomes in PAH.

Methods

Modeling methodology and use of data is summarized in Figure 4.1. Further details on each step of model development are described below.

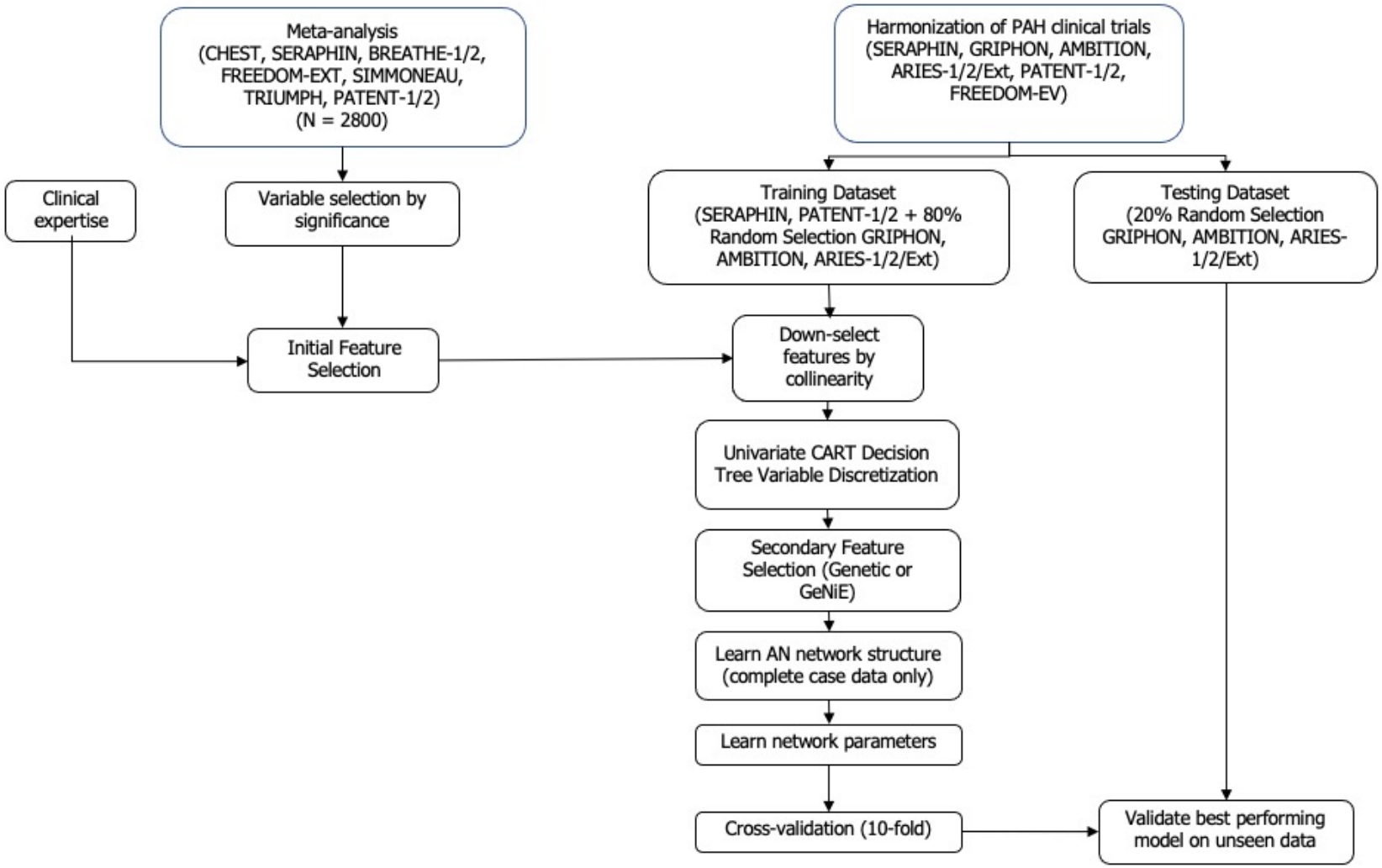


Figure 4.1: Methodology for Bayesian Network Model Learning.

Harmonizing Multiple Clinical Trial Datasets

To obtain enough training dataset for a machine learning model, we collaborated directly with the U.S. Food and Drug Administration to create a hi-fidelity harmonized dataset using multiple clinical trials conducted on pulmonary arterial hypertension patients since 2004. This source of data provides a rich set of clinical variables for individual-level patients at baseline as well as their clinical outcomes, as well as international representation from multiple clinical sites. This harmonized dataset was comprised of six clinical trials conducted from 2004-2019 individual patient level data: A Study of First-Line Ambrisentan and Tadalafil Combination Therapy in Subjects With Pulmonary Arterial Hypertension (AMBITION)(42), Ambrisentan in Patients With Moderate to Severe Pulmonary Arterial Hypertension with its open-label extension (ARIES-1/2/E)(14; 108), Phase III Clinical Worsening Study of UT-15C in Subjects With PAH Receiving Background Oral Monotherapy (FREEDOM-EV)(109), Selexipag in Pulmonary Arterial Hypertension (GRIPHON)(37), A Study to Evaluate Efficacy and Safety of Oral BAY63-2521 in Patients With Pulmonary Arterial Hypertension with its open-label extension (PATENT-1/2)(21; 110), and Study of Macitentan on Morbidity and Mortality in Patients With Symptomatic Pulmonary Arterial Hypertension (SERAPHIN)(16). All clinical trials were published previously. Only patients with known vital status at the end of one year from baseline (dead or censor day >= 365) were included in the training dataset.

Meta-analysis and Initial Feature Selection

Studies have shown that initial feature selection for statistical and machine learning models conducted in a unique dataset can avoid overfitting and increase effective sample

size.(111; 112) Clinical variables were assessed in a distinct group of clinical trials: Bosentan therapy for pulmonary arterial hypertension (BREATHE-1)(15), A Study to Evaluate Efficacy and Safety of Oral BAY63-2521 in Patients With Chronic Thromboembolic Pulmonary Hypertension (CHEST)(113), An Open-Label Extension Trial of UT-15C Sustained-release (SR) in Subjects With Pulmonary Arterial Hypertension (FREEDOM-EXT)(114), Continuous subcutaneous infusion of treprostinil, a prostacyclin analogue, in patients with pulmonary arterial hypertension: a double-blind, randomized, placebo-controlled trial (SIMMONEAU)(32), and Clinical Investigation Into Inhaled Treprostinil Sodium in Patients With Severe Pulmonary Arterial Hypertension (TRIUMPH)(36). PATENT-1/2 and SERAPHIN were also included in the meta-analysis to provide additional data on hemodynamics and NTproBNP/BNP, which were not measured in any other trial in the meta-analysis. Hence, neither SERAPHIN nor PATENT are used in the hold-out dataset, to prevent “memory-leakage” between feature selection and final hold-out testing (i.e. training and validating on the same subjects).

Using a two-step method, univariate Cox proportional hazards were first measured in each respective trial for prediction of one-year survival, then Stouffer’s method was applied for p-value aggregation from each individual trial to determine a final relative variable importance. Due to the increased chances in finding clinical variables with a significant p-value in univariate analyses, Bonferroni’s correction was applied to control the family-wise error rate of univariate analysis on multiple clinical variables (FWER) to be < 0.05. Only variables with an adjusted p-value of < 0.05 were considered candidates for the model, unless PAH clinical experts recommended their candidacy based on

literature review and significance was limited by data availability. All aggregated p-values, adjusted p-values, and data availability for features in meta-analysis are reported.

As many variables are strongly correlated, a correlation heatmap was produced and variables with Pearson's correlations > 0.6 were down-selected, favoring the better predictor of survival (lower p-value from the meta-analysis). For example, if stroke volume and stroke volume index had a correlation of 0.9, but stroke volume index had a smaller p-value in the meta-analysis, stroke volume index was chosen over stroke volume.

Univariate Decision Tree Discretization

Classification and Regression Trees (CART) decision trees were used to select and pre-process features for the discrete Augmented Naïve Bayesian network. The benefit of using one machine learning model to inform another is that the initial method allowed for a data-driven process for discretizing continuous variables (a necessary pre-processing step for discrete Bayesian networks) by pre-learning their relationship with the outcome to be used in the network. Multiple publications have demonstrated that discretization of variables can improve generalization and performance of machine learning models by preventing overfitting of variables with threshold effects on the predicted outcome.(81; 115) For example, there is a suspected saturation limit to NTproBNP (in excess of 3000 pg/mL) where higher values are no longer informative. Specifically, CART decision trees were chosen to optimize (i.e. minimize) the negative Brier score, in which is considered a proper scoring method, as discussed in Chapter 1. These univariate decision trees were then 10-fold cross-validated within the training dataset to determine the optimal number of cut-points for a single variable, while reducing

overfitting, optimizing the negative Brier score. Once the number of cut-points was established, decision trees were refit to all training data to produce final cut-points. Groups separated by a cut-point that did not have a significant difference in survival rate, as determined from a two-sided Barnard's exact test, were re-merged, a technique referred to as "pruning".(116)

Advanced Feature Selection with Differential Evolution versus GeNiE methods

A novel method for determining candidate feature groups was developed, using a hybrid "filter" and "wrapper" method with differential evolution. Differential evolution is meta-heuristic approach to obtain a globally optimized solution for non-convex objective functions.(77) The objective function was defined with two specific goals: 1) to maximize the information obtained from features after discretization, 2) reduce redundancy between features.

The first goal could be satisfied simply by accounting for all variables. However, a greater number of features also increases propensity for model overfitting. Therefore, the use of the smallest feature set that still maximizes performance is desired. The pairwise redundancy of features was determined through Kendall's tau for agreement in ranked outcomes (i.e. concordance). Discretized features that ranked patients identically would have a Kendall's tau of one or 100% concordance; conversely, discretized features that ranked patients in the exact opposite manner would have a Kendall's tau of negative one or -100% concordance. As features were discretized on a univariate basis, it's advantageous to include in pairs of features that encode different types of information regarding clinical outcomes. For example, if men are found to be at higher risk of one-

year mortality than women, but so are patients with connective tissue disease, which disproportionately affects women, its advantageous to consider both features. Conversely, for features that have very similar agreement in ranking patients, the features likely demonstrate a correlative physiological effect, and only the one that best ranks the patients with the true outcome (also determined by Kendall's tau) should be prioritized. For example, if six-minute walk distance ranks patients nearly the same as their New York Heart Association Functional Class, which is partially causally based on six-minute walk distance, but six-minute walk distance betters agrees with the true outcome, NYHA should be dropped in favor of six minute walk distance.

Finding a group of features that optimizes this balance is conducted with differential evolution, which creates an initial "population" of feature groups, finds the groups with the best "fitness" for the objective function (i.e. sum of concordance between each feature and the outcome), and then "evolves" these groups towards a final feature set. Multiple rounds of differential evolution were conducted with different rates of "penalization" of feature redundancy (i.e. sum of concordance between how each feature pair ranks patients, where features that rank patients more similarly give a greater penalty and are likely to be dropped). Higher penalization results in groups with fewer features, lower penalization results in groups with more features. Lastly, the objective function included a term to penalize fractional values to drive the optimization towards choosing or dropping features, rather than finding solutions with partial feature consideration.

Figure 4.2 shows the final pseudocode for differential evolution feature selection.

function CALCULATEOBJECTIVEFUNCTION(parameterCoefficients, redundancyPenaltySize)
-select features with parameterCoefficient > 0.9 for each selected discretized feature
-calculate concordance between feature and outcome
-sum concordances for all selected features
-for each feature pair, calculate concordance of their respective ranking (i.e. redundancy)
-force parameterCoefficient values to zero or one by penalizing fraction remainders
-objective function = redundancyPenaltySize(sum of concordances between feature pairs) - (sum of concordances between feature and outcome) + fraction penalty
-return objective function value
end function

function DIFFERENTIALEVOLUTION(calculateObjectiveFunction, parameterBounds, maxiterations = 100)
-set relative redundancy penalty equal to the constant redundacyPenaltySize
-set all parameterBounds to [0,1]
-generate initial population size of parameterCoefficients with Latin Hypercube Sampling
-calculate objective function value with function calculateObjectiveFunction(parameterCoefficients, redundancyPenaltySize)
-choose best candidate parameterCoefficients that minimize the objective function
-evolve candidates towards global maximum
-repeat for maximum 100 iterations or until objective function value no longer changes by a small threshold
-return best parameterCoefficient solution
-select features with parameterCoefficients > 0.9
end function

Figure 4.2: Pseudocode for Differential Evolution Feature Selection.

The method is compared, using cross-validation, to multiple other feature selection methods for Bayesian networks: 1) selection of top ten features with maximal mutual information with the outcome, 2) selection of features based on causal modeling (aka greedy thick-thinning conducted in GeNiE), 3) selection of features using GeNiE PC, 4) selection of features using the GeNiE structure learning method for augmented Naïve Bayes.

Augmented Naïve Bayesian Network

The harmonized clinical trial dataset described above was used to build an augmented naïve Bayesian network. Bayesian networks aim to maximize the joint probability of a structure by learning probabilistic relationships between "parent" and "child" nodes. All nodes that are not linked are conditionally independent of all others given the outcome. Augmented naïve Bayesian networks are more flexible than both Tree-Augmented Naïve (TAN) Bayesian networks, in that they allow for interactions between variables but do not force assumptions about the number of "parents", whereas TAN requires that every node has at least one parent beyond the outcome, regardless of strength of linkage. As the differential evolution feature selection method specifically aims to remove redundant variables, fewer nodes will have a strong "parent" node, allowing for the "naïve" assumption of conditional independence among variables to better hold.

GeNie modeling software was used to build multiple network classifiers with different numbers of variables, based on the outputs of the differential evolution feature selection with increasing penalty. Only a subset of the training data is used for the training process (20%), as it requires complete records to learn interactions. Parameters of the model were then improved by using expectation-maximization, which iteratively improves the likelihood of the parameters using the full training dataset, including patients with missing variables. Ten-fold cross validation on each candidate model was used to test different subsets of the candidate variables, repeated ten times with a different random seed to determine folds. The best final model was determined by choosing the model with the fewest variables but no significant drop in performance (i.e. cross-validation AUC) compared to the model with the most variables. Differences in AUC were determined with

a nonparametric Kruskal-Wallis test, with a *post hoc* Dunn test for multiple comparisons, and $p < 0.05$ was used to determine significance. Final model performance was measured on the hold-out validation set and compared to several other published PAH risk calculators, including COMPERA(59), French (FPHR)(60), REVEAL 2.0(61) and PHORA [1.0](65).

Results

Patient Population Summary

Summary statistics for the harmonized clinical trial dataset is provided in Table 4.1. This reflects the population with no early censored patients (i.e. all patients with no death event were observed for at least a year from baseline). Overall, the harmonized dataset reflected the common demographics of Group I PAH – majority female (78.5%), non-Hispanic white (63.6%), idiopathic PAH (57.3%) or PAH associated with connective tissue disease (29.78%), with an average age of 48 years and an "overweight" BMI (average 26.4 kg/m^2). Patients were also most commonly NYHA Class II or III (47.70% and 50.08%, respectively) and were on one or more PAH medications prior to randomization to treatment (63.70%).

Harmonized Clinical Trial Dataset							
	AMBITION	**SERAPHIN**	**GRIPHON**	**PATENT-1/2**	**ARIES-1/2**	**FREEDOM-EV**	**Total**
Total Patients	N = 443	N = 720	N = 1060	N = 237	N = 113	N = 584	3157
Placebo/Treatment	223/ 220	480/240	518/542	178/59	89/24	284/300	1385 P/1772 T
Gender	342/101	551/169	848/212	192/45	88/25	458/126	2479 F/678 M
Age	54 ± 14	46 ± 16	48 ± 15	49 ± 16	52 ± 15	45 ±15	48 ± 16
Race							
White (Non-Hispanic)	395	389	685	141	87	311	2008
Black or African American	31	19	25	2	5	14	96
Asian	10	203	232	89	3	256	793
Other	6	109	118	5	18	3	259
Body Mass Index	27.8 ± 6.5	25.5 ± 5.9	26.7 ± 6.2	26.0 ± 5.7	26.5 ± 5.4	26.0 ± 6.0	26.4 ± 6.1
NYHA							
I	0	1	7	5	2	19	34
II	132	379	485	100	42	368	1506
III	311	326	557	130	61	196	1581
IV	0	14	11	2	8	1	36
Six Minute Walk Distance	350.5	360.2	352.8	366.7	343.3	397.0	363.1± 90.9
Etiology of PAH							
Connective Tissue	167	217	308	56	38	154	940
Congenital Heart Disease	6	60	103	19	0	37	225
IPAH	238	399	598	144	69	361	1809
Familial PAH	11	13	22	6	0	9	61
HIV	8	10	7	0	5	7	37
Toxin	13	21	22	2	1	16	75
Portal	0	0	0	10	0	0	10
Treatment Naive	443	258	210	122	94	0	1127/2011
Deaths in One-Year	18	49	95	16	21	32	231 (7.3%)

Table 4.1: Summary Statistics for Harmonized Clinical Trial Dataset

Feature Selection Meta-Analysis

Results from meta-analysis for variables determined to be significant (adjusted p-value < 0.05) are shown in Table 4.2. Variables are ranked in order of lowest p-value (first) to highest p-value (last). Table 4.3 includes all other variables that were not significant (adjusted p-value > 0.05).

Variable	Total N	Aggregated p-value	Bonferroni Correction
Total Bilirubin	2539	3.91E-10	3.55E-08
Six Minute Walk Distance	1430	1.09E-08	9.89E-07
NYHA FC II	1852	1.16E-08	1.06E-06
Mixed Venous Oxygen Saturation	1739	1.50E-08	1.37E-06
Stroke Volume Index	1937	2.07E-08	1.89E-06
Sex: Female	1435	2.84E-08	2.59E-06
Heart Rate	1769	4.23E-08	3.85E-06
Left Heart Stroke Work	1827	6.37E-08	5.79E-06
Pulmonary Arterial Elastance	1649	1.14E-07	1.04E-05
Cardiac Power Output	2753	1.36E-07	1.24E-05
Stroke Volume	1546	3.42E-07	3.11E-05
Race: Asian	2048	9.16E-07	8.34E-05
Body Surface Area	1397	1.01E-06	9.21E-05
Total Peripheral Resistance	1891	2.11E-06	1.92E-04
Alkaline Phosphatase	2492	4.93E-06	4.48E-04
NYHA FC IV	992	5.20E-06	4.73E-04
Cardiac Output	2003	5.65E-06	5.14E-04
Creatinine	1560	8.08E-06	7.36E-04
Cardiac Efficiency	1622	8.72E-06	7.94E-04
Aspartate Aminotransferase	1539	1.11E-05	1.01E-03
Pulmonary Vascular Resistance	1654	1.38E-05	1.26E-03
Use of Diuretics	992	2.14E-05	1.95E-03
PAH Medication (Treatment Naïve)	2038	2.58E-05	2.35E-03
Neutrophil	422	3.12E-05	2.84E-03
Blood Urea Nitrogen	1946	3.14E-05	2.86E-03
Body Mass Index	1780	4.78E-05	4.35E-03
Lactate Dehydrogenase	1492	5.33E-05	4.85E-03
Cardiac Index	1060	5.64E-05	5.14E-03
Sodium	1791	5.66E-05	5.15E-03
Albumin	1710	7.58E-05	6.90E-03
Chloride	1793	8.87E-05	8.10E-03
History: Right Heart Failure	1734	9.46E-05	8.61E-03
NYHA FC III	860	1.00E-04	9.10E-03
Neutrophil: Lymphocyte	211	1.00E-04	9.10E-03
Right Atrial Pressure	1501	3.41E-04	3.10E-02
Right Ventricular Stroke Work	1534	4.16E-04	3.78E-02

Table 4.2: Results of Meta-Analysis for Significant Variables.

Variable	Total N	Aggregated p-value	Bonferroni Correction
Diastolic Pulmonary Arterial Pressure	1514	7.35E-04	6.69E-02
Right Ventricular Stroke Work Index	1529	1.02E-03	9.30E-02
Hematocrit	2214	1.09E-03	9.95E-02
Age	2414	1.21E-03	1.10E-01
Leukocyte	211	1.40E-03	1.27E-01
Respiration Rate	994	2.14E-03	1.94E-01
Systolic Blood Pressure	1342	2.31E-03	2.10E-01
Race: Hispanic	860	2.40E-03	2.18E-01
Urate	105	3.00E-03	2.73E-01
Race: Caucasian	860	3.40E-03	3.09E-01
Chronic Thromboembolism Pulmonary Hypertension	155	4.00E-03	3.64E-01
Pulmonary Vascular Resistance Index	616	4.20E-03	3.82E-01
BORG (Perceived Exertion during Six Minute Walk)	1766	4.40E-03	4.00E-01
NT-Pro BNP	146	5.00E-03	4.55E-01
Platelet	1909	5.08E-03	4.62E-01
Aspirin	992	6.63E-03	6.03E-01
Creatinine Clearance	382	7.32E-03	6.66E-01
Right Ventricular Cardiac Work	865	7.89E-03	7.18E-01
Mean Pulmonary Arterial Pressure	1519	8.29E-03	7.55E-01
Race: Native American	819	8.30E-03	7.55E-01
Gamma Glutamyl Transferase	251	8.55E-03	7.78E-01
Time since PAH	821	1.01E-02	9.19E-01
Right Atrial to Pulmonary Capillary Wedge Pressure	621	1.35E-02	1.23E+00
Estimated glomerular Filtration Rate	942	1.35E-02	1.23E+00
Bilirubin Direct	730	1.51E-02	1.37E+00
Triacylglycerol Lipase	423	1.67E-02	1.52E+00
Lymphocyte	211	1.76E-02	1.60E+00
History: Raynaud	287	1.90E-02	1.73E+00
Use of Statin	109	1.90E-02	1.73E+00
Pseudocholinesterase	146	2.70E-02	2.46E+00
Right Ventricular Cardiac Work Index	860	2.82E-02	2.57E+00
Pulmonary Arterial Compliance	642	2.84E-02	2.58E+00
Etiology: Congenital Heart Disease	954	3.07E-02	2.80E+00
Red Blood Cell Count	809	3.23E-02	2.94E+00
Monocyte	211	3.54E-02	3.22E+00
Diastolic Blood Pressure	1141	3.97E-02	3.62E+00
Systolic Pulmonary Arterial Pressure	849	4.28E-02	3.89E+00
Oxygen Supplement	109	4.70E-02	4.28E+00
Erythrocyte	704	5.18E-02	4.71E+00
Urea	146	5.60E-02	5.10E+00
Etiology: CTD	954	5.67E-02	5.16E+00
Pulmonary Artery Pulsatile Index	653	5.78E-02	5.26E+00
EQ-5 (Quality of Life Score)	285	5.90E-02	5.37E+00
Etiology: Idiopathic Pulmonary Hypertension	109	6.70E-02	6.10E+00
Hemoglobin	211	7.41E-02	6.74E+00
Acute Hemodynamic Index	205	7.82E-02	7.12E+00
Creatine Kinase	146	8.00E-02	7.28E+00
Use of Digoxin	109	8.30E-02	7.55E+00
Pulmonary Capillary Wedge Pressure	1642	9.18E-02	8.36E+00
Ratio Monocyte To Leukocyte	146	9.70E-02	8.83E+00
Ratio Pulmonary Vascular Resistance to Systemic Vascular Resistance	596	1.68E-01	1.53E+01

Pulmonary Hemodynamic Index	653	2.46E-01	2.23E+01
Mean Arterial Pressure	210	3.79E-01	3.45E+01

Table 4.3: Results of Meta-Analysis for Insignificant Variables

As determined from the correlation heatmap, features were assessed for high collinearity. The full correlation heatmap is shown in Figure 4.3.

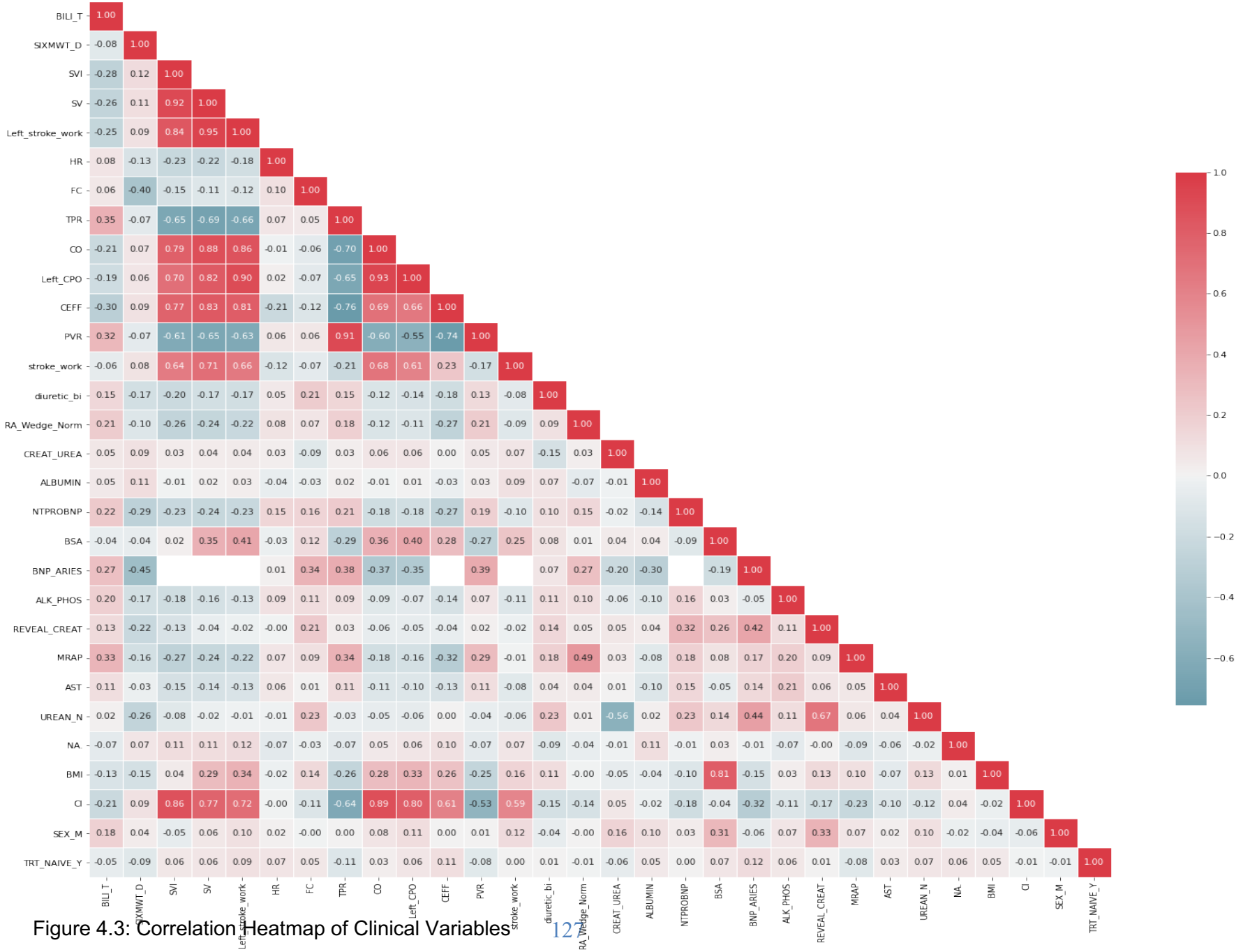


Figure 4.3: Correlation Heatmap of Clinical Variables

Specifically, left stroke work, stroke volume, cardiac power output, total pulmonary resistance, cardiac output, cardiac efficiency, pulmonary vascular resistance, cardiac index and right ventricular stroke work were downselected due to high correlations with stroke volume index. BMI was downselected due to high correlation with body surface area.

The following variables were not available in the validation dataset and were dropped from feature candidacy: mixed venous oxygen saturation, pulmonary arterial elastance, lactate dehydrogenase, serum chloride, history of right heart failure, ratio of neutrophil to lymphocyte, neutrophil cell count. From expert opinion and literature review, age, NTproBNP/BNP, ratio of right atrial pressure to pulmonary capillary wedge pressure, and presence of connective tissue disorder were additionally considered for feature candidacy despite non-significant p-values in the meta-analysis. The final feature selection prior to model training is shown in Table 4.4.

Variables Included from Meta-analysis	Variables Included from Expert Opinion
Total Bilirubin	Age
Six Minute Walk Distance	NTproBNP/BNP
NYHA Functional Class	Right Atrial Pressure to PCWP
Stroke Volume Index	Presence Connective Tissue Disorder
Sex (Male/Female)	
Heart Rate (Sitting)	
Body Surface Area	
Alkaline Phosphatase	
Creatinine	
Aspartate Aminotransferase	
Use of diuretics	
PAH Medication (Treatment Naïve)	
Sodium (Serum)	
Albumin	
Mean Right Atrial Pressure	
Blood Urea Nitrogen	

Table 4.4. Candidate Features Following Meta-Analysis and Expert Opinion

CART Decision Tree Pre-processing and Significance Testing

All continuous features shown in Table 4.4 were assessed for discretization using univariate CART decision trees. An example of a learned tree is shown in Figure 4.4, for the variable total bilirubin.

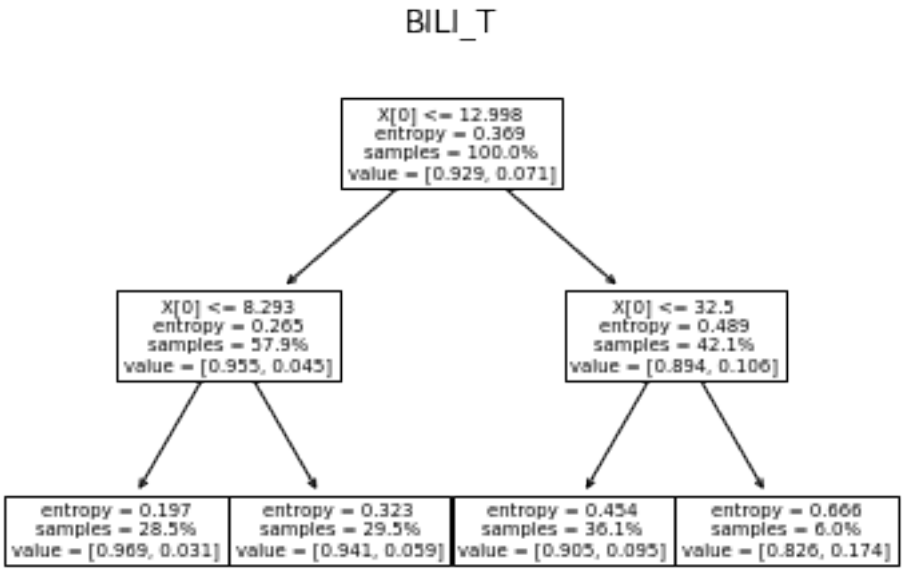


Figure 4.4. Example of Univariate Decision Tree to Discretize Continuous Clinical Variables (ex. Serum Total Bilirubin)

A table of all learned cut-points for all candidate features is shown in Table 4.5, including the estimated survival for each group above or below the threshold and p-values of the Barnard's exact test between groups, with corrections for multiple comparisons.

Variable	Identified Cut-points	Effect on Prognosis	Estimated Survival Rate	p-value (Barnard's Exact Test)
Total Bilirubin (µmol/L)	Very Low <= 8.3 8.3 < Low <= 13 13 < Intermediate <= 32.5 High > 32.5	Improves Neutral Worsens Significantly Worsens	96.9% 94.1% 90.5% 82.6%	Very Low vs Low: 0.044 Low vs Intermediate: 0.044 Intermediate vs High: 0.044
Six Minute Walk Distance (m)	High > 346.5 244.5 < Intermediate <= 346.5 Low <= 244.5	Improves Worsens Significantly Worsens	96.6% 90.3% 78.1%	Low vs Intermediate: 0.0002 Intermediate vs High: 0.0008
Albumin (U/L)	High > 43.5 36.5 < Intermediate <= 43.5 Low <= 36.5	Improves Worsens Significantly Worsens	95.8% 91.7% 84.1%	Low vs Intermediate: 0.037 Intermediate vs High: 0.008
Stroke Volume Index (mL)	High > 3.6 Low <= 3.6	Improves Worsens	98.5% 90.1%	Low vs High: 0.00013725
Heart Rate (bpm)	Low <= 83.5 83.5 < Intermediate <= 89.5 High > 89.5	Improves Worsens Significantly Worsens	95.1% 92.2% 85.2%	Low vs Intermediate: 0.026 Intermediate vs High: 0.012
RAP:PCWP	Low <= 1.087 High > 1.087	Improves Significantly Worsens	95.0% 89.4%	Low vs High: 0.00161055
NTproBNP (ng/L)	Low <= 533.5 533.5 < Intermediate <= 1834 High > 1834	Improves Neutral Significantly Worsens	98.0% 94.8% 81.0%	Low vs Intermediate: 0.028 Intermediate vs High: 3.31e-6
Alkaline Phosphatase (U/L)	Low <= 122.5 High > 122.5	Neutral Significantly Worsens	94.1% 86.0%	Low vs High: 0.00037799
Creatinine (µmol/L)	Low <= 77.05 Intermediate <= 107.85 High > 107.85	Improves Neutral Significantly Worsens	95.4% 92.1% 83.7%	Low vs Intermediate: 0.024 Intermediate vs High: 0.006
Aspartate Aminotransferase (U/L)	Low <= 22.5 High > 22.5	Improves Worsens	95.2% 90.9%	Low vs High: 0.0018
Sodium (mmol/L)	Low <= 137.5 High > 137.5	Neutral Significantly Worsens	93.8% 89.8%	Low vs High: 0.029

Age (yrs)	Low <= 27.5 27.5 < Intermediate <= 64.5 High > 64.5	Significantly Worsens Neutral Significantly Worsens	89% 94.2% 90%	Low vs Intermediate: 0.027 Intermediate vs High: 0.027
Body Surface Area (m^2)	High > 1.82 1.74 < Intermediate <= 1.82 1.6 < Low <= 1.74 Very Low <= 1.6	Improves Significantly Worsens Neutral Significantly Worsens	95.1% 88.4% 94.7% 90.7%	Very Low vs Low: 0.032 Low vs Intermediate: 0.032 Intermediate vs High: 0.013
BUN (mmol/L)	Low <= 7.755 High > 7.755	Neutral Significantly Worsens	94.5% 87.6%	Low vs High: 0.0017
Connective Tissue Disorder	Y N	Worsens Neutral		Yes vs No: 0.02
NYHA FC	I/II III IV	Improves Worsens Significantly Worsens		I vs II: p = 0.37 II vs III: $p = 4.53\times10^{-3}$ III vs IV: $p = 6.46\times10^{-3}$
Use of Diuretics	Y N	Worsens Neutral		Yes vs No: 0.028
Sex	M F	Worsens Neutral		Male vs Female: 0.0018

Table 4.5. Results of Univariate Decision Tree Analysis.

No cut-points could be identified using mean right atrial pressure that would create groups with significant differences in one-year survival rate. Results of Barnard's exact testing for discrete variables is shown in Table 4.5, with correction for multiple comparisons.

Results of Differential Evolution Feature Selection

Figure 4.5 shows the concordance of patient ranking (from low to high risk based on estimated probability of survival) between each variable, using Kendall's tau.

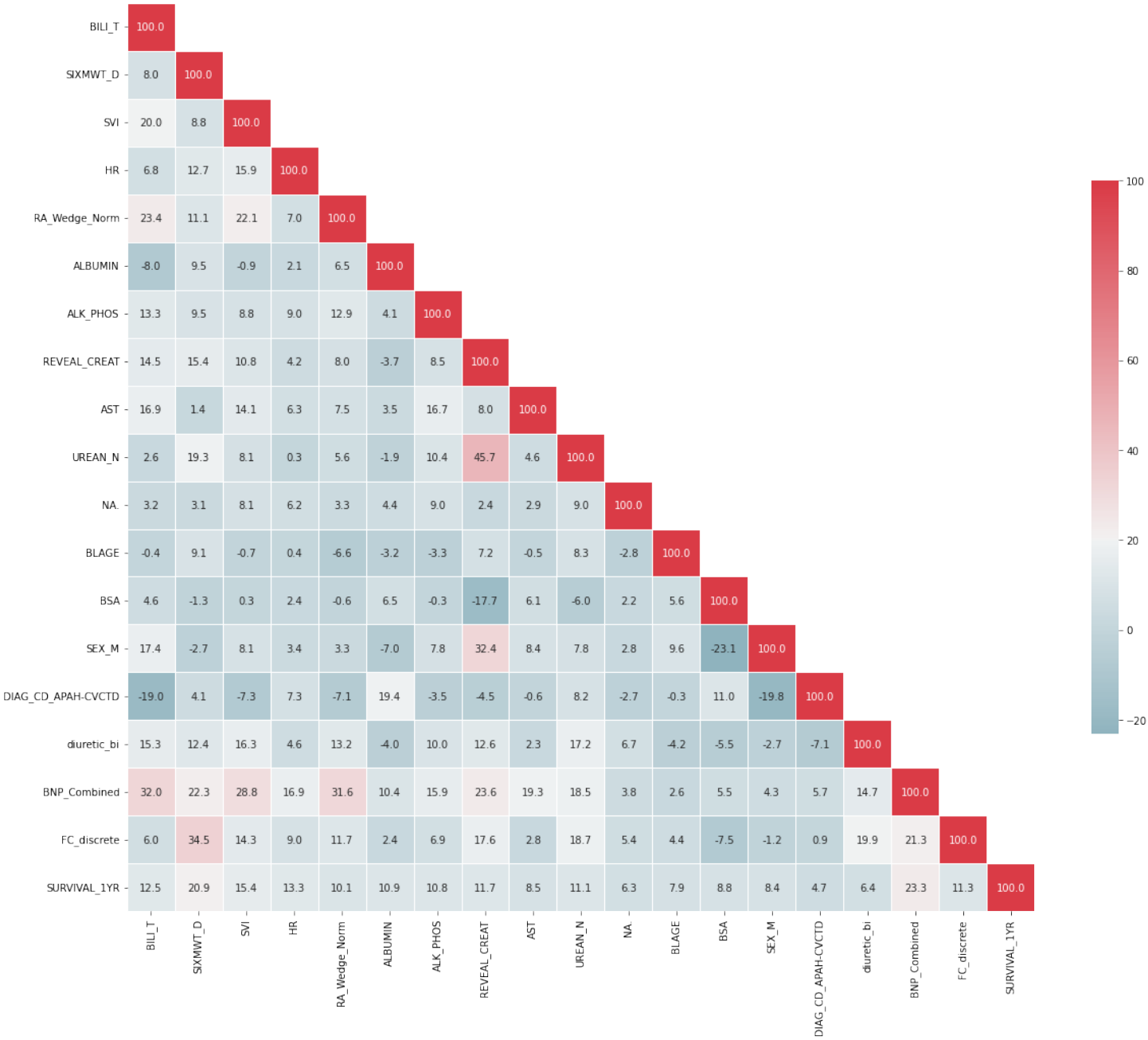


Figure 4.5. Kendall's tau (concordance) of outcome prediction by clinical variables

Variables with highest concordance with the outcome are strong univariate predictors of survival. The top three variables with highest concordance are six-minute walk distance (20.9%), NTproBNP/BNP (23.3%), and stroke volume index (15.4%).

Other highly positive relationships between variables demonstrate a potentially significant degree of redundancy in how they rank patients. For example, total bilirubin and NTproBNP/BNP have a concordance of 32.0%, meaning that patients with low survival predicted by a “high” NTproBNP also tend to have “high” total bilirubin. The top three highest concordances (i.e. greatest redundancy) were seen between serum creatinine and blood urea nitrogen, six-minute walk distance and NYHA functional class, and male sex and serum creatinine.

Conversely, negatively concordant variables rank patients in opposite directions. For example, patients with low survival predicted by connective tissue disorder are less likely to be male, but males are predicted to have worse survival than females (concordance -19.8%). The lowest concordances (i.e. least redundancy) were seen between male sex and body surface area, male sex and presence of connective tissue disorder, and total bilirubin and presence of connective tissue disorder. Interestingly, connective tissue disorder had a negative or low concordance with nearly all clinical variables, suggesting that most connective tissue disorder patients, despite having a higher risk of mortality, do not typically also have other high-risk features.

Results from the differential evolution feature analysis are shown in Table 4.6

Model Number	Penalty Size	6MWD	HR	Age	BSA	Sex	CTD	ALB	AP	BILI	BUN	SVI	BNP	NA	Di	Cr	AST	FC	R:P
1	0	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	0.05	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	
3	0.075	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓
4	0.10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					
5	0.15	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓					
6	0.175	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓							
7	0.2	✓	✓	✓	✓	✓	✓	✓					✓	✓	✓				

Table 4.6: Differential Evolution Feature Selection

Penalty size decreases the number of features selected, but also affected the variables chosen, based on the degree of redundancy with other variables that may be present.

Bayesian Networks and Final Model Performance

Figure 4.6 shows the cross-validation results for all models using a group of features identified by differential evolution, as well as the GeNiE selection methods.

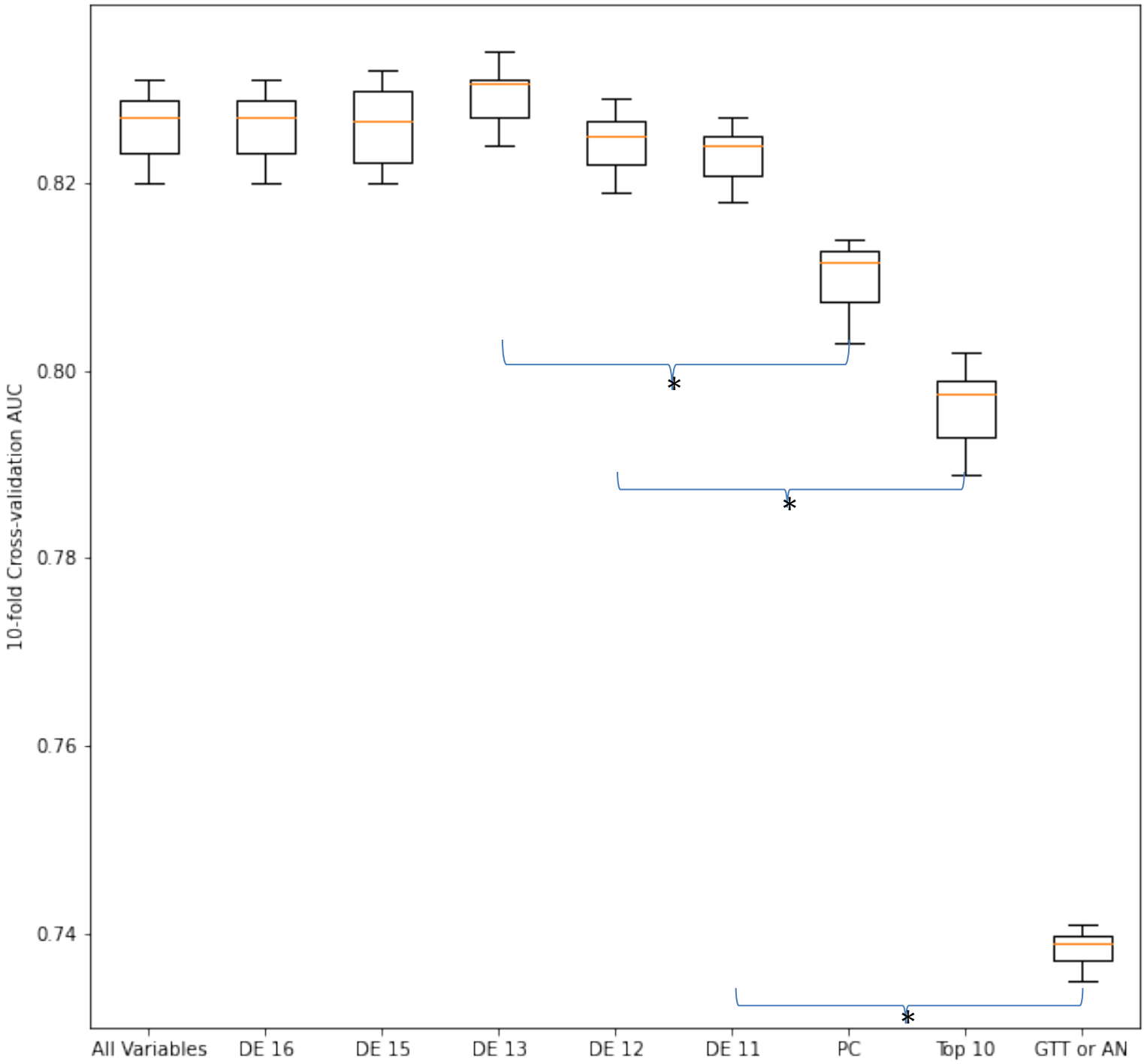


Figure 4.6: Comparison of Model Performance for Different Feature Selection Methods

.

As shown, all models using differential evolution feature selection outperformed the GeNiE selection methods, although the differences were only significant down to a penalty size of 0.10 (13 variables). Of the GeNiE feature selection methods, PC performed the best, identifying a model that only required 8 variables. Feature selection through greedy thick-thinning and augmented naïve Bayesian feature selection both only identified three variables for prediction: NTproBNP, alkaline phosphatase, and albumin. The classifier built with these three variables was significantly worse than all differential evolution models, but not significantly worse than the models built with PC nor top 10 independent variable feature selection. Differential evolution Model #4 (13 variables) had the highest average cross-validation AUC, but the drop in performance was not significantly different between Model #4 and Model #6 (11 variables). Therefore, Model 6 was chosen as the final model for validation in the holdout set, shown in Figure 4.7.

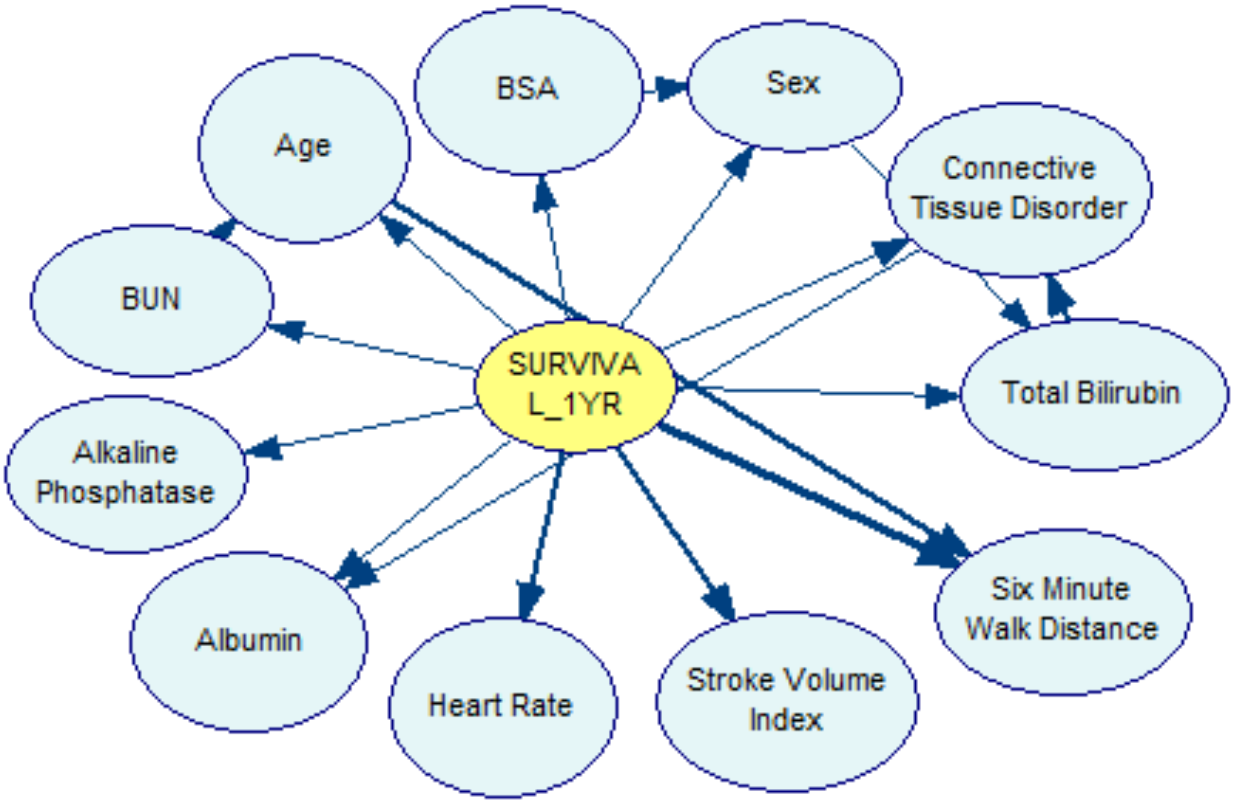


Figure 4.7: Final Network for PHORA 2.0, Using Differential Evolution Feature Selection (Model #6)

Figure 4.8 shows the receiver-operator curve for the final chosen model versus published PAH risk calculators. The model was statistically significantly better than the next best AUC (REVEAL 2.0).

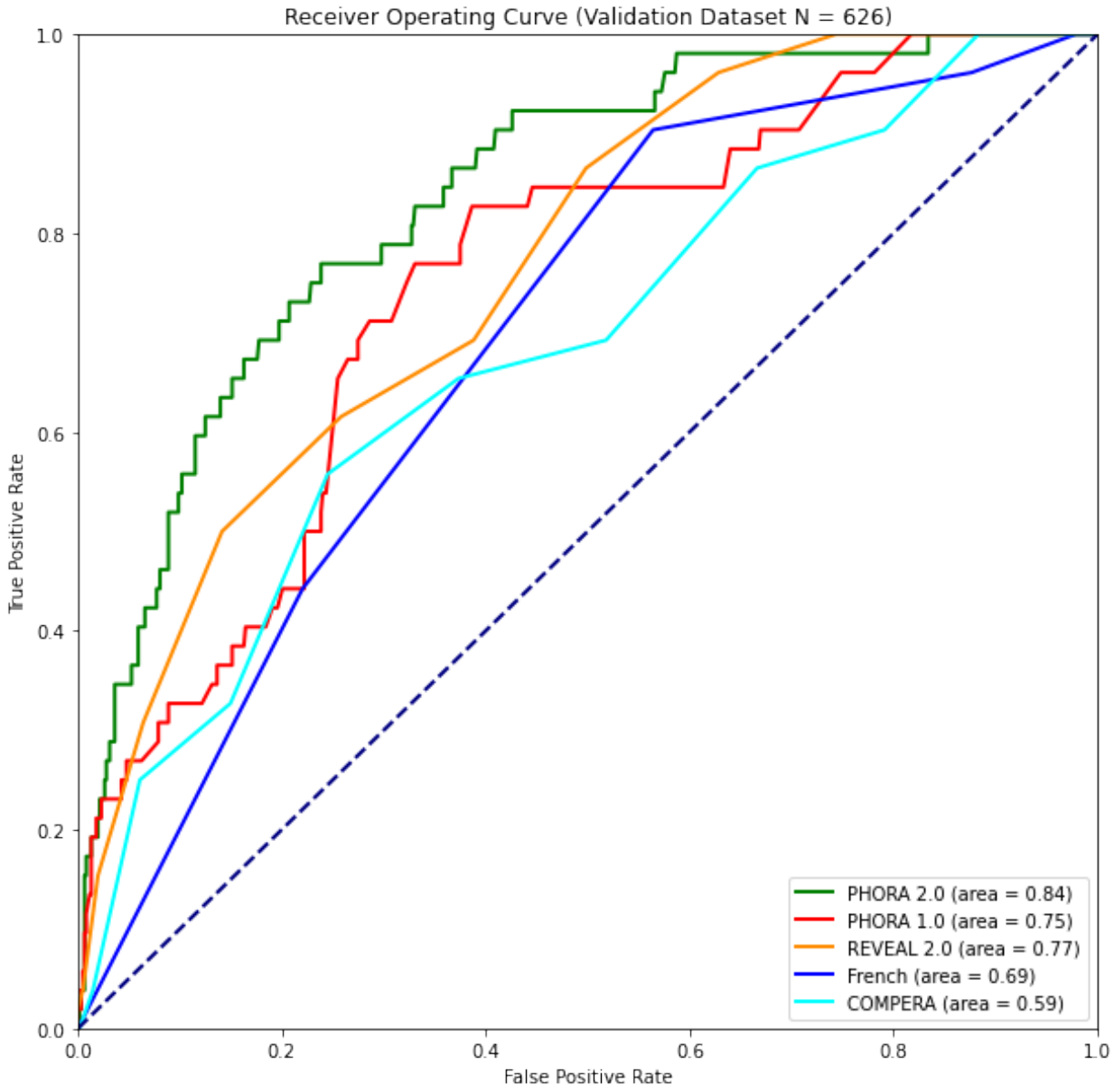


Figure 4.8: Validation on Hold-Out Set for New Model (PHORA 2.0) versus All Risk Calculators.

Discussion

This study demonstrates the potential power of machine learning for improving prognostic tools for pulmonary arterial hypertension. It has generated the first machine learning model built with data-driven techniques to generate excellent performance (AUC > 0.80) on hold-out data. Further, it has allowed for new hypothesis generation

into the pathophysiology of PAH, specifically regarding systemic consequences of right heart failure and integration of novel biomarkers.

Risk stratification is one of the most critical clinical tools for the treatment and clinical study of pulmonary arterial hypertension. As patients' treatment strategy depends on accurate clinical risk assessment, improving prognostic tools is paramount to improving PAH clinical decision science. As shown in Chapters 2 and 3, PAH clinical decision science can go much further with more accurate prognostic tools, allowing for both clinical trial enrichment and risk-benefit assessment in treatment strategy.

The modeling methodology conducted in this chapter highlights many significant clinical takeaways about PAH risk stratification in general. Based on the meta-analysis, stroke volume index was the most important hemodynamic value for assessing survival. These results contrast to some degree with current 2015 European Respiratory Society guidelines to use cardiac index as the primary measure of systemic perfusion. This highlights that there may be a difference in cardiac index achieved through high sympathetic nervous activation (i.e. high resting heart rate) versus a cardiac index achieved through greater RV ejected volume. As demonstrated from the meta-analysis and decision tree analysis, a higher resting heart rate confers worse survival. Thus, clinicians should consider decoupling cardiac index to stroke volume and heart rate when evaluating right heart function in patients with PAH.

Although discretizing features can lead to a loss of information between the predictor variables and the outcome, it is still a highly preferred method in clinical decision science, significantly easing the model learning process, and can prevent overfitting in some cases. Effective cut-point analysis requires a cross-validated means

of choosing hyperparameters, such as the number of cut-points to identify, to further ensure against overfitting, as well as pruning methods. Of clinical interest, decision tree analysis for NTproBNP identified cutpoints (low risk <= 533.5, 533.5 < intermediate risk <= 1834, high > 1834) within 25% of the values chosen for the 2015 ERS guidelines (low risk <= 400, 400 < intermediate risk < 1400, high > 1400). In contrast, cut-points for six minute walk distance (low risk > 346, 346 >= intermediate risk > 244.5, high risk <= 244.5) were less similar, within 48% of the value chosen for the ERS guidelines (low risk > 440, 165 < intermediate risk <= 440, high risk < 165). Cut-points used in this novel model for six-minute walk distance were less extreme in value than those used in the guidelines, yet were still able to confer significant differences in one-year survival rate between all groups. This suggests that the guidelines are too extreme in their low and high values for six-minute walk distance.

Significance testing did not demonstrate a significant difference in survival rate between NYHA Functional Class I and Class II patients, yet there was a significant difference between Class III and IV. Interestingly, however, Class I and Class IV were roughly just as frequent in the dataset. Therefore, this difference is not explained by sample size alone. Although the Class I designation may predict other significant clinical outcomes, such as quality of life, it does not appear to be a significant predictor of mortality outcomes. This supports the ERS guidelines, which specifies that both Class I and II are given the same risk level designation.

The 2015 ERS guidelines recommends regular laboratory workups for PAH patients that includes total bilirubin, liver aminotransferases (though specifies alanine aminotransferase only for patients on ERA therapy), creatinine, and serum sodium for

patients with PAH every 3-6 months.(10) However, there is no mention in the guidelines on how to determine what values in each of these labs are signs of clinical worsening. Hyperbilirubinemia (defined as a total serum bilirubin of > 20.5 μmol/L) and hyponatremia (defined as serum sodium <137 mmol/L) are considered additional evidence in evaluation for placement on the lung transplant waiting list, yet no current risk model accounts for either clinical value. This model is the first means of risk stratification to account for these systemic consequences of right heart failure. Decision tree analysis for serum sodium agreed almost exactly (high risk serum sodium < 137.5 mmol/L) with the definition of hyponatremia. Decision tree analysis for total bilirubin demonstrated that the variable is more informative than a single value, identifying three groups with significantly different survival rates, and showing a significant decrease in chances of survival for total bilirubin values as low as 13 μmol/L.

A recent review article has discussed the importance and effect of right ventricular failure on organ systems. In the kidneys, venous congestion and hypoperfusion cause chronic congestive nephropathy and low perfusion renal injury, respectively.(103) The consequences of chronic congestive nephropathy are renal edema, increased interstitial pressures, and tubular compression, all of which act to reduce glomerular filtration. Further, the activation of inflammatory mediators caused by backflow results in glomerular and interstitial damage, fibrosis, and reduced kidney function, resulting in salt and water retention and proteinuria. This results in elevated creatinine and blood urea nitrogen levels. Low perfusion renal injury results in tissue fibrosis, cell necrosis, and impaired cell signaling. As the kidneys play a key role in blood pressure and fluid homeostasis, their dysfunction can cause a vicious cycle that

further strains the right ventricle, pushing it closer to failure. Increased circulating inflammatory factors caused by kidney injury also can worsen pulmonary hypertension directly. Patients with repeated subclinical kidney injuries are at risk for developing chronic kidney disease.(103) In the liver, again, venous congestion and hypoperfusion cause chronic congestive injury (hepatopathy) and ischemic injury (hepatitis). These both result in a loss of function for the liver, decreasing its synthetic function (reduced albumin), increasing hepatocellular injury, and result in liver fibrosis. While ischemic hepatitis is rare and has clear, severe symptoms, assessing which patients are at greater risk for developing such a condition is crucial to prevention. Chronic congestive hepatopathy can remain subclinical and is believed to be a significant risk factor of later developing ischemic hepatitis. Despite the clear clinical understanding of these issues, current risk stratification methods largely overlook kidney and liver biomarkers. In reality, these two organs may serve as the “canaries in the coalmine”, especially as hemodynamics alone have been repeatedly shown to be insufficient in risk stratifying patients. This machine learning model is the first guidance to identify optimized cut-points for identifying patients at higher risk based on blood urea nitrogen, serum sodium, total bilirubin, alkaline phosphatase, and aspartate aminotransferase, as well as show the interactions between variables.

While the harmonization of clinical trials allowed for a large enough dataset to develop a more accurate machine learning model, there remains a crucial need for large, opensource data to both validate and improve upon PAH risk stratification. Data access and privacy remains a major challenge for all machine learning applications to clinical data. Though some researchers have begun to recognize this need and create

publicly funded data registries for PAH (namely, the Pulmonary Hypertension Association Registry or PHAR), these registries require greater resources for consistent data collection of crucial biomarkers, hemodynamics, and imaging.(117) Further, as genetics, epigenetics, and proteomics become further implicated in the pathophysiology of PAH, bigger data still will be required to make the necessary discoveries into PAH genotypes and phenotypes to finally arrive at curative treatments.

Overall, machine learning has proven itself as a successful tool for knowledge discovery and succinct synthesis of large datasets, in addition to a means of more accurately risk stratifying PAH patients. Lessons from this model demonstrated the clear importance of overlooked liver and kidney biomarkers, as well as the importance of inclusion of connective tissue etiology to reduce underestimation of risk within this PAH subpopulation.

Limitations

This study had a few limitations. External validation of the machine learning model in a contemporary (2019 or later) dataset would be ideal. While there are many registry datasets currently used in international PAH research, access to these datasets is extremely limited. This remains a major limitation in clinical decision support research in general.

The training dataset used is now considered sub-contemporary, as it includes patient baseline data from as early as 2004. Clinical practice and management of PAH has changed significantly since that time period, which can affect baseline rates of one-year mortality as well as the relevance of certain clinical variables.

Interactions between systemic consequences of right heart failure, such as subclinical liver injury, and pharmaceutical treatment, could be crucial to understanding PAH mortality. Specifically, with the increased clinical use of double or triple oral therapies, it's unclear if a potentially compromised liver can metabolize an increasing number of medications. Another possibility is that patients compromised liver function are more susceptible to drug-induced liver injury that results in medication withdrawal, and subsequent clinical worsening.

Modeling treatment interactions, biomarkers, and functional capacity (symptoms and exercise capacity) is likely necessary to achieving higher performance of PAH risk models. This increasing number of variables will likely require more complex and sophisticated modeling paradigms (deep neural networks) as well as even larger datasets. Probably and Approximately Correct Model theory maintains that these sample sizes can be estimated by the VC dimension of a model, by which a deep neural network has a VC dimensions several times larger than an augmented naïve Bayesian network (depending on hyperparameters). However, empirical studies have shown there is something of a paradox in that deep neural networks can still generalize well to unseen data with significantly fewer examples than required by PAC theory. Further still, the use of embedded domain knowledge may significantly reduce required sample sizes for machine learning models by restricting the number of learned relationships and parameters required. As will be discussed in the last chapter of this dissertation, novel physiological models of PAH that demonstrate how perfusion and congestion reduce right ventricular work capacity may be the key to creating more accurate PAH risk models.

Conclusion

In this chapter, Bayesian network modeling, first introduced with the development of the PHORA model, has been further optimized using differential evolution feature selection, allowing for significant gains in risk prediction accuracy. Further, the application of this machine learning technique has further supported the importance of liver and kidney biomarkers, as well as incorporation of demographics and etiology such as sex, age, and presence of connective tissue disorder in the assessment of risk. External validation of these models in large datasets is required before implementation into clinical practice but allows for well-specified hypothesis generation into further studying the systemic consequences of right ventricular failure.

Chapter 5 : Energetic Model of Right Ventricular Failure

Introduction

Right ventricular (RV) function is the primary determinant of survival in patients with pulmonary arterial hypertension (PAH). [103] Abnormal increases in right ventricular afterload lead to eventual decreases in cardiac output and increased venous congestion, a vicious cycle that further strains the right ventricle and ends in right ventricular failure.

Per the 2015 European Respiratory Society guidelines, clinicians are instructed to focus directly on metrics of RV output (RV ejection fraction, cardiac output, stroke volume, etc) as markers of disease severity and RV adaptation.[10] These metrics are crucial; however, less consideration is given to how RV output is a direct function of RV metabolic substrate delivery, particularly the oxygen required to power ATP production to fuel RV contraction.(118) As PAH progresses, multiple factors can limit right coronary perfusion including increased RV systolic pressure that drastically reduces right coronary systolic flow and reduction in cardiac output that reduces the driving pressure for coronary perfusion. [119] Further, the estimated prevalence of coronary artery disease in PAH patients is 28%, a condition that further limits coronary perfusion.[120] The culmination of these factors leads to a reduction in total oxygen delivered to the right ventricle, and reduced right coronary perfusion reserve has been observed in PAH patients.(121) Once right ventricular oxygen extraction and coronary flow reserves are exhausted, patients will

be less capable of tolerating a new insult or demand upon their cardiovascular system, such as physical exertion.

The RV requires high oxygen delivery to overcome the high afterload, but it must also be able to use this oxygen efficiently. Myocardial external efficiency, the observed ratio of stroke work to RV myocardial oxygen consumption, has been shown to be higher in PAH patients of New York Heart Association Class II (NYHA II) versus Class III/IV, linking inefficient oxygen use to disease severity.[122] However, it is unclear if this inefficiency reflects higher RV load, a preferential switch to aerobic glycolysis over oxidative phosphorylation, or one of the other possible causes that have been identified, including but not limited to septal bowing, neurohumoral metabolism, and increased heart rate. A statistical model published in 2011 by *Wong et al*. suggested that the main predictors of RV myocardial oxygen consumption (MVO_2) in patients with PAH were heart rate and systolic pulmonary arterial pressure but did not provide the specific physiological mechanism by which these values determine MVO_2. [123] As discussed in Chapter 1, pressure-volume area (PVA) models of ventricular energetics have been applied to other heart failure modalities but only represent a fraction of oxygen consumption (roughly 40%), are difficult to measure reliably, and are more invasive as they require manual manipulation of venous return.(87; 124) Further, they do not reflect true thermodynamic first principles of energy expenditure under isovolumic contraction.(87) Although pressure-volume loops are not currently used to assess PAH patients, a more accurate energetic model of the RV under high afterload could prove valuable to assessment of

PAH RV dysfunction and guiding optimization of extracorporeal membrane oxygenation, which can be used as a bridge-to-transplant (10)'.

In order to improve understanding of how right ventricular oxygen demand and efficiency changes due to PAH, a novel right ventricular work model is proposed that builds upon the earlier work of *Elbeery et al* on the left ventricle.[87] The underlying premise of this model is that (1) ventricular total work capacity is dictated primarily by its oxygen supply and that (2) energy wasted due to high "isovolumic" work demands results in less energy available to produce a sufficient cardiac output. Here, the model is modified to more accurately portray the relationship between right ventricular hemodynamics and oxygen consumption, resulting in the first model to use thermodynamic first principles to calculate right ventricular power output specific to PAH. The accuracy of the original and adjusted models was then examined using data from clinical studies of patients with idiopathic PAH and from large animal models that underwent a combination of acute pulmonary hypertension with or without hypoxia. Finally, the clinical relevance of this model for improving assessment of RV function is discussed.

Methods

Clinical versus Animal Study Model Differences

The metabolic model of RV function relies on a variety of hemodynamic and blood oxygenation data. Due to discrepancies in available instrumentation and avoidance of invasive measures in clinical studies, some modeling approaches differ

between animal study data and clinical study data. A summary of differences is provided in Table 5.1 and discussed in more detail in the section below.

	CLinical studies (N= 15)	Animal Studies (N=6)
cause of increased pulmonary impedance	Idiopathic	Pulmonary Arterial Banding
Right coronary Flow measurement	PET Imaging	Ultrasonic flow probe
pulmonary arterial flow measurement	Estimated with invented flow profile	Ultrasonic flow probe
right ventricular oxygen extraction fraction measurement	PET Imaging	Estimated from literature values

Table 5.1: Summary of Differences Between Modeling Approach for Clinical and Animal Studies

Metabolic Right Ventricular Function Model

The model presented here builds upon the model of *Elbeery et al*, originally designed and validated in canine left ventricles.[87] Changes are made to make the model more appropriate for PAH pathophysiology, RV function, and human clinical studies.

In a normal adult human heart, oxidative phosphorylation is responsible for almost all ATP generation in myocardial tissue (95%), with glycolysis providing the remaining percentage.(125) While typically the work demands on the RV are low and oxygen supply is abundant, in situations of high afterload, there is significant evidence that RV work is limited by RV oxygen supply, similar to that of a healthy left ventricle.(121; 123) This model, therefore, uses this assumption to couple RV oxygen consumption directly to its energy expenditure via the following equation:

$$Total\ Mechanical\ Energy\ (J) = 20.2 \times O_2\ Consumption, \quad (10)$$

where *O_2 Consumption* is in mL per heartbeat and the constant factor is based on the theoretical efficiency of oxidative phosphorylation with mixed substrates (i.e. 20.2 joules per mL of oxygen).(125) The units for Equation 10 are joules (J) per heartbeat. In this equation, the small contribution of anaerobic energy sources (glycolysis) are considered to be negligible (see Discussion for further comment).-where *O_2 Consumption* is in mL per heartbeat and the constant factor is based on the theoretical efficiency of oxidative phosphorylation with mixed substrates (i.e. 20.2 joules per mL of oxygen).(125)

The rate of oxygen consumption can, therefore, be coupled to the power output of the RV, or work per unit time.

$$20.2 \times O_2\ Consumption\ Rate\ \left(\frac{J}{min}\right) = Total\ Mechanical\ Energy \times HR, \quad (11)$$

where *HR* is heart rate (beats/min) and *O_2 Consumption Rate* is in volume per min (mL/min). The units for Equation 11 are joules per minute (J/min)

The rate of O_2 delivery to the right ventricle free wall is determined primarily by right coronary flow in humans.(126) In sheep, there is anatomical evidence that the left anterior descending branch of the left coronary artery also supplies blood to the upper region of the right ventricle, but its contribution has not been quantified, and this model assumes its effect on right ventricular oxygen supply to be negligible.(127) This relationship is modeled as:

$$O_2\ Delivery\ Rate\ \left(\frac{mL}{min}\right) = 1.34 \times OEF \times Q_{RCA} \times S_a \times C_{hb}, \quad (12)$$

where S_a is arterial oxygen saturation (%), *OEF* is RV oxygen extraction fraction, Q_{RCA} is right coronary flow (mL/min), C_{hb} is the concentration of hemoglobin (g/mL), and the constant (1.34) is the hemoglobin oxygen carrying capacity (1.34 mL of O_2 per gram of hemoglobin).(128) The units for Equation 12 are milliliters per minute (mL/min). To avoid invasively measuring arteriovenous differences in oxygen saturation during acute sheep studies, RV *OEF* was not directly measured, but instead estimated to be 50% in normoxic and 60% in hypoxic conditions

based on previous studies.(129; 130) RV *OEF* for clinical studies, however, was determined directly through PET imaging.[122],[128],[131]

According to the *Elbeery et al* model, total mechanical energy (*TME*), in joules (J), is dictated by both an internal index of heat (referred to here as isovolumic work) and stroke work(87)*:*

$$TME\ (J) = Isovolumic\ Work + Stroke\ Work, \qquad (13)$$

The units for Equation 13 are joules (J). RV total power output (*TPO*), in joules per minute (J/min), is therefore the sum of isovolumic work and stroke work over a cardiac cycle:

$$TPO\ \left(\frac{J}{min}\right) = HR\ \times (Isovolumic\ Work + Stroke\ Work), \qquad (14)$$

The units for Equation 14 are joules per minute (J/min).

In the *Elbeery et al* model, isovolumic power is estimated via:

$$Isovolumic\ Power\ (J/min) = HR \times mRV_{EP} \times (RV_{EDV} - V_0) \times (1.33 \times 10^{-4}), \qquad (15)$$

where mRV_{EP} is the mean right ventricular ejection pressure (mmHg) and RV_{EDV} is right ventricular end-diastolic volume during systole.(87) The units for Equation 15 are joules per minute (J/min). The constant (1.33×10^{-4}) is the conversion factor from mmHg-mL to joules, and V_0, the dead volume that is determined by the ventricular end-systolic pressure-volume relationship, is considered to be negligible.(87)

Ventricular stroke work is calculated as:

$$\text{Stroke Work (J/min)} = \int_0^T P_{PA} Q_{PA}\ , \qquad (16)$$

where P_{PA} and Q_{PA} are the instantaneous pulmonary arterial pressure and flow measurements, respectively, and *T* is the period of RV ejection. The units for Equation 16 are joules per minute (J/min).

Although Q_{PA} could be directly measured in the acute sheep studies via flow probe, this cannot be done for clinical patients and pulmonary arterial flow was instead approximated using an invented flow profile.(132) Although Q_{PA} could be directly measured in the acute sheep

studies via flow probe, this cannot be done for clinical patients and pulmonary arterial flow was instead approximated using an invented flow profile.(132) The invented flow profile was based on the following piecewise function:

$$Q_i\ (mL/min) = \begin{cases} A(SV)\ exp(-3t)\ sin\left(\frac{\pi t}{t_{es}}\right) & 0 < t < t_{es} + 0.02 \\ Q_i(t_{es} + 0.02)\left[1 - \frac{t - t_{es} - 0.02}{0.02}\right] & t_{es} + 0.02 < t < t_{es} + 0.04 \\ 0 & t_{es} + 0.04 < t < t_{ed} \end{cases} \quad (17)$$

Where *A* is an amplitude proportionality factor, *SV* is the stroke volume in mL, t_{es} is the time at which RV ejection ends in seconds, t_{ed} is the time at which RV diastole ends in seconds, and 0.04s is assumed to be the duration of pulmonary valve regurgitation.(132) The units for Equation 17 are milliliters per minute (mL/min). The value of the amplitude proportionality factor *A* was calibrated such that mean $Q_i(t)$ was equivalent to each patient's respective cardiac output. Finally, the invented flow ($Q_i(t)$) and discrete pulmonary arterial pressure data were integrated for the first 5 sequential heartbeats during the period of ejection using the trapezoidal method to calculate stroke work. The stroke work for each heartbeat was divided by the period of ejection (*T*) to obtain stroke power.

$$\text{Stroke Power (J/min)} = \frac{1}{T}\int_0^T P_{PA} \times Q_{PA} \quad (18)$$

The units for Equation 18 are joules per minute (J/min). Stroke power is averaged over five heartbeats to get a steady-state measurement at rest. Details on how the period of ejection was determined in the RV pressure waveform can be found in the Appendix.

The following adjusted model was compared to the *Elbeery et al* model(87):

1) Assume that isovolumic work is proportional to raising the RV end-systolic volume (RV_{ESV}) to the mean ejection pressure, rather than the end-diastolic volume (RV_{EDV}). This provides a lower bound for total RV work, by assuming that the majority of work is performed in raising the entire end-diastolic volume to the mean ejection pressure and that stroke volume ejection is a passive consequence of the resulting pressure gradient

between the RV and distal PA. Mathematically, this can be demonstrated by decomposing RV_{EDV} into two components for Equation 15 (recall that V_0 is considered negligible):

$$Isovolumic\ Power\ (J/min) = HR \times mRV_{EP} \times (RV_{ESV} - SV) \times (1.33 \times 10^{-4}), \quad (19)$$

Then, the assumption is made that ejecting the stroke volume is negligible to isovolumic power:

$$Isovolumic\ Power\ (J/min) = HR \times mRV_{EP} \times (RV_{ESV}) \times (1.33 \times 10^{-4}), \quad (20)$$

2) Assume that isovolumic work is proportional to the difference between mean ejection pressure (mRV_{EP}) and end-diastolic pressure (RV_{EDP}), rather than mean ejection pressure alone. Arguably, initially raising the right ventricle to its end-diastolic pressure is work performed by the left ventricle. In patients with PAH, RV_{EDP} can be significantly raised due to venous congestion, making this change crucial to model accuracy.

With the modifications above, the adjusted model changes Equation 15 to be:

$$Isovolumic\ Power\ (J/min) = HR \times (mRV_{EP} - RV_{EDP}) \times (RV_{ESV}) \times (1.33 \times 10^{-4}) \quad (21)$$

The units for Equation 21 are joules per minute (J/min).

3) In humans, zeroth harmonic stroke power (the product of mean pulmonary arterial pressure and cardiac output) can be calculated directly. Out of necessity, the non-zeroth harmonic stroke power is estimated using the method of *Sauoti et al*.(133), by which oscillatory work is estimated to be proportional to pulmonary arterial pulse pressure.

Stroke work is therefore estimated in the adjusted model by:

$$Stroke\ Power\ (J/min) = HR \times (mPAP) \times (SV) \times (1.33 \times 10^{-4}) + (0.156 \times PP) \quad (22)$$

The units for Equation 22 are joules per min (J/min).

The most accurate means for calculating stroke work is provided in Equation 18 and used when applying the adjusted model to animal data. Equation 22 is only used as a stopgap to estimate pulsatile work without the ability to measure Q_{PA} in humans.

Total power output for the *Elbeery et al* model is then:

$$TPO\ \left(\frac{J}{min}\right) = HR \times mRV_{EP} \times (RV_{EDV} - V_0) \times (1.33 \times 10^{-4}) + \frac{1}{T}\int_0^T P_{PA} * Q_{PA}, \quad (23)$$

which is the sum of Equation 15 and Equation 16.(87) The units for Equation 23 are joules per minute (J/min)

This is contrasted with total power output for the adjusted model:

$$TPO\ (J/min) = HR \times (mRV_{EP} - RV_{EDP}) \times (RV_{ESV}) \times (1.33 \times 10^{-4}) + HR \times (mPAP) \times (SV) \times (1.33 \times 10^{-4}) + 0.156 \times PP, \quad (24)$$

which is the sum of Equation 21 and Equation 22. The units for Equation 24 are joules per minute (J/min).

The final RV efficiency metric is then:

$$RV\ Efficiency\ (\%) = \frac{Useful\ Stroke\ Power}{Total\ Power\ Output}, \quad (25)$$

where *Useful Stroke Power* (J/min) is defined as non-pulsatile stroke power, in order to differentiate between stroke power that represents forward flow into the pulmonary artery versus stroke power loss to pulsatile effects that does not generate forward blood flow.

$$Useful\ Stroke\ Power\ (J/min) = mPAP \times CO, \quad (26)$$

and total power output is defined by either Equation 23 for the *Elbeery et al* model or Eq. 24 for the adjusted model. The units for Eq. 26 are joules per minute (J/min).

Animal Studies

Surgical Methods and Instrumentation

The validity of the model was tested first using acute studies with adult, Dorset breed male sheep (*Ovis aries*) (N=6, 60 ± 3 kg). All sheep received humane care in compliance with the "Guide for the Care and Use of Laboratory Animals" and all methods were approved by the University of Michigan Committee for the Use and Care of Animals. Anesthesia was induced with 6-9 mL/kg of propofol and maintained with 1-3% inhaled isoflurane (Abbot Laboratories, Chicago, IL). Sheep were mechanically ventilated at all times with 100% oxygen set to a tidal volume of 10 mL/kg. The respiratory rate was adjusted to maintain an arterial PCO_2 ($PaCO_2$)

between 35 and 45 mmHg. Lastly, a carotid arterial line and left jugular venous line were placed and then connected to fluid coupled pressure transducers (ICU Medical, Inc., San Clemente, CA) for animal management.

A left thoracotomy was performed, and a 24 mm perivascular flow probe (24AX, Transonic Systems, Inc., Ithaca, NY) was placed around the main pulmonary artery to measure cardiac output (CO) continuously with a T400 Flowmeter (Transonic Systems; Ithaca, NY). Umbilical tape was passed around the main pulmonary artery and a Rummel tourniquet was used to increase pulmonary resistance when needed. Right coronary artery dissection was then performed by retracting the right atrium laterally to expose the atrio-ventricular groove. Blunt dissection with electrocauterization was performed to isolate a small section of the right coronary artery and place a 4-mm perivascular flow probe (4AX, Transonic Systems, Inc., Ithaca, NY) to measure right coronary artery flow rate continuously (T400 Flowmeter, Transonic Systems; Ithaca, NY).

Proximal PA and RV pressures were measured continuously via 14G angiocatheters (Becton, Dickinson and Company, Franklin Lakes, NJ) connected to fluid coupled pressure transducers (ICU Medical, Inc., San Clemente, CA). Lastly, an 8F pulmonary artery catheter (model 777F8, Edwards Lifesciences, Irvine, CA) was introduced into the right jugular vein and its tip placed in the pulmonary artery distal to the flow probe and umbilical tape. An ECG lead was placed on each of the sheep's legs and attached to a patient monitor (Marquette Solar, Marquette, WI). The monitor was then interfaced with the Vigilance CEDV monitor (Edwards Lifesciences, Irvine CA) to measure heart rate (*HR*), right ventricular ejection fraction (*EF*) and end diastolic volume (RV_{EDV}).

Animal Experimental Methods

A baseline data set was taken after completing instrumentation. Proximal PA, RV, and central venous pressures and PA flow rate were acquired digitally at 250 Hz for seven seconds using LabVIEW (National Instruments, Austin, TX). All other data, specifically HR, RV end diastolic volume, arterial saturation and hemoglobin concentration (measured with Radiometer Blood Gas Analyzer, Brea, CA), were hand recorded. Thereafter, RV afterload and oxygen saturation to the RV were varied. A hypoxic condition was induced by lowering the inspired oxygen fraction, FIO_2, and thus arterial oxyhemoglobin saturation (S_a). The afterload was increased by adjusting the pulmonary blood flow zeroth harmonic impedance modulus, Z_0, using a Rummel tourniquet on the main PA. This index has been a preferred index for afterload by some physiologists because, unlike mPAP alone, it continues to rise as pulmonary vascular resistance (PVR) is increased to very high levels and, unlike PVR, it includes the load placed on the RV by left atrial pressure. Experimental groups were defined by the change in Z_0 applied over baseline values, ΔZ_0, and S_a. The following four groups were thus examined; (i) Normal afterload, normoxia (n=5): ΔZ_0 = 0 mmHg/(L/min), S_a = 100%; (ii) High afterload, normoxia (n=6): ΔZ_0 = 4 mmHg/(L/min), S_a = 100%; (iii) Normal afterload, hypoxia (n=6): ΔZ_0 = 0 mmHg/(L/min), S_a = 75%; and (iv) High afterload, hypoxia (n=6): ΔZ_0 = 4 mmHg/(L/min), S_a = 75%. Each condition was maintained for four hours. All data were taken every 40 minutes as at baseline. Both S_a and Z_0 were assessed every 20 minutes and adjusted as needed to maintain their target values. After four hours, the sheep were euthanized with 90-150 mg/kg of IV pentobarbital (Fatal-Plus, Vortech Pharmaceuticals, Dearborn, MI).

Clinical Studies

Through collaboration with the Pulmonary Hypertension Group at the Vrije Universiteit (VU) University Medical Center in Amsterdam, Netherlands, RV hemodynamic and metabolic

data for 15 idiopathic pulmonary artery hypertension (IPAH) patients were analyzed. Of the 15 patients (aged 26-71 years), 14 were female and 1 was male. The protocol was approved by the Medical Ethics Review Committee of VU University Medical Center. Each patient gave written informed consent before the study. Each patient was grouped according to the New York Heart Association Classification (NYHA). RV hemodynamic data was obtained through cardiac Magnetic Resonance Imaging (MRI), cardiac Positron Emission Tomography (PET), and right heart catheterization. Details on calculation of oxygen extraction fraction are provided in *Wong et al*.(122)two-minute mark from a total five-minute acquisition time. (122) Briefly, the entire acquisition time for PET imaging was 10 minutes, using 40 frames. Oxygen extraction fractions were determined from volume-weighted average time-activity curves of the identified right ventricular wall region, identified through standard methods of a $C^{15}O$-PET image and cardiac MRI. Pulmonary arterial pressure, right ventricular pressure, and ECG data was recorded simultaneously for most patients (n=11) and asynchronously for those remaining (n=4) during baseline in a supine position at a sampling rate of 1 kHz for a period of 5 minutes. Values are ensemble averaged per heartbeat over five heartbeats. For hemodynamics values and volumes, ensemble averages were calculated over five heartbeats, extracted starting from the A summary of all provided measurements are given in Results section. Further details of the full clinical study can be found in *Wong et al.* (122)

Statistical Analysis

To validate both models (*Elbeery et al*'s model and the adjusted model), the coefficient of determination (R^2) was measured between the RV O_2 consumption and RV total power output for all clinical data; for sheep, data was analyzed at baseline and after 4-hours (when hemodynamics were considered stable). Statistical significance (p-value) of each model's coefficient of determination and the 95% confidence intervals were calculated. The 95%

confidence interval for each regression coefficient and y-intercept term were also determined. The regression coefficient is expected to reflect the theoretical efficiency of oxidative phosphorylation (20.2 J/mL O_2) and the y-intercept term is expected to be non-significant, reflecting the negligible contribution of anaerobic sources to total power (i.e. power output in the absence of oxygen).

To examine the relationship between RV efficiency is related to oxygen consumption rate, the coefficient of determination between these two variables and its statistical significance are measured. RV efficiency is hypothesized to be negatively related to oxygen consumption rate (i.e. high RV efficiency allows for low RV oxygen consumption and vice versa).

To examine the relationship between RV oxygen consumption and percentage of type of work, average and standard error of the mean of percent contribution of isovolumic power, useful stroke power, and pulsatile stroke work to total work are reported by class for NYHA Class II and Class III/IV patients.

Finally, to examine the relationship between RV efficiency, oxygen consumption, and symptoms of right heart failure, a MANOVA analysis was used to show statistically significant differences between NYHA Class II and III/IV patients in RV efficiency from both the *Elbeery et al* and the adjusted model as well as significant differences in oxygen consumption rate. NYHA Class III and IV patients were combined to ensure sufficient statistical power. In terms of disease severity, both NYHA Class III and IV are considered to have significant limitations in functional capacity versus NYHA Class II, making the grouping appropriate.(55) Given a null hypothesis of no difference in these three metrics between low and high NYHA class, the alternative hypothesis is that higher NYHA class (more severe disease state) would have lower RV efficiency as calculated by either model and higher RV oxygen consumption rates. *Post hoc* analyses with Bonferroni correction were used to determine specifically which dependent

variables (mechanical efficiency from the *Elbeery et al* model, mechanical efficiency from the adjusted model, and/or oxygen consumption rate) were significantly different between Class II and III/IV.

Results

Table 5.2 shows the summary of hemodynamics and demographics for the clinical studies. Briefly, patients were evenly split between low (NYHA Class II) and high (NYHA Class III and IV) disease severity.

	NYHA II (n=8)	NYHA III (n=4)	NYHA IV (n=3)
Mean PA Pressure (mmHg) *mPAP*	47.4±11.7	58.8±13.3	65.3±22.9
Cardiac Output (L/min) *CO*	5.6±0.9	4.2±1.1	3.5±1.4
Ejection Fraction *EF*	46±13.3	31±3.5	19±4.4
Right Coronary Mean Blood Flow (mL/min) *Rc MBF*	47.5±15.5	64.6±20.8	55.2±12.6
Oxygen Extraction Fraction *OEF*	0.6±0.15	0.7±0.16	0.9±0.08
RV Free Wall End Diastolic Mass (g)	72.5±22.7	85.1±19.7	88.1±12.9
Age	43.6±14.2	48.3±16.8	42.0±10.8
Sex (Female/Male)	7/1	4/0	3/0

Table 5.2. Summary of Hemodynamics and Demographics for Clinical Studies.

Average values for hemodynamics worsened with disease severity (higher afterload and reduced cardiac output and RV ejection fraction), but there was significant overlap between all values across disease severity. Interestingly, average right coronary mean blood flow was larger

for NYHA Class III than Class II but was lower for NYHA Class IV than Class III. This demonstrates compensation in Class III, but decompensation in Class IV due to the effect of reduced cardiac output on coronary perfusion and RV oxygen delivery.

Figure 5.1 shows the linear relationship between O_2 consumption (mL/min) and right ventricular total power output (J/min) for clinical studies using both the *Elbeery et al* model (Figure 5.1a) and the newly proposed model (Figure 5.1b).

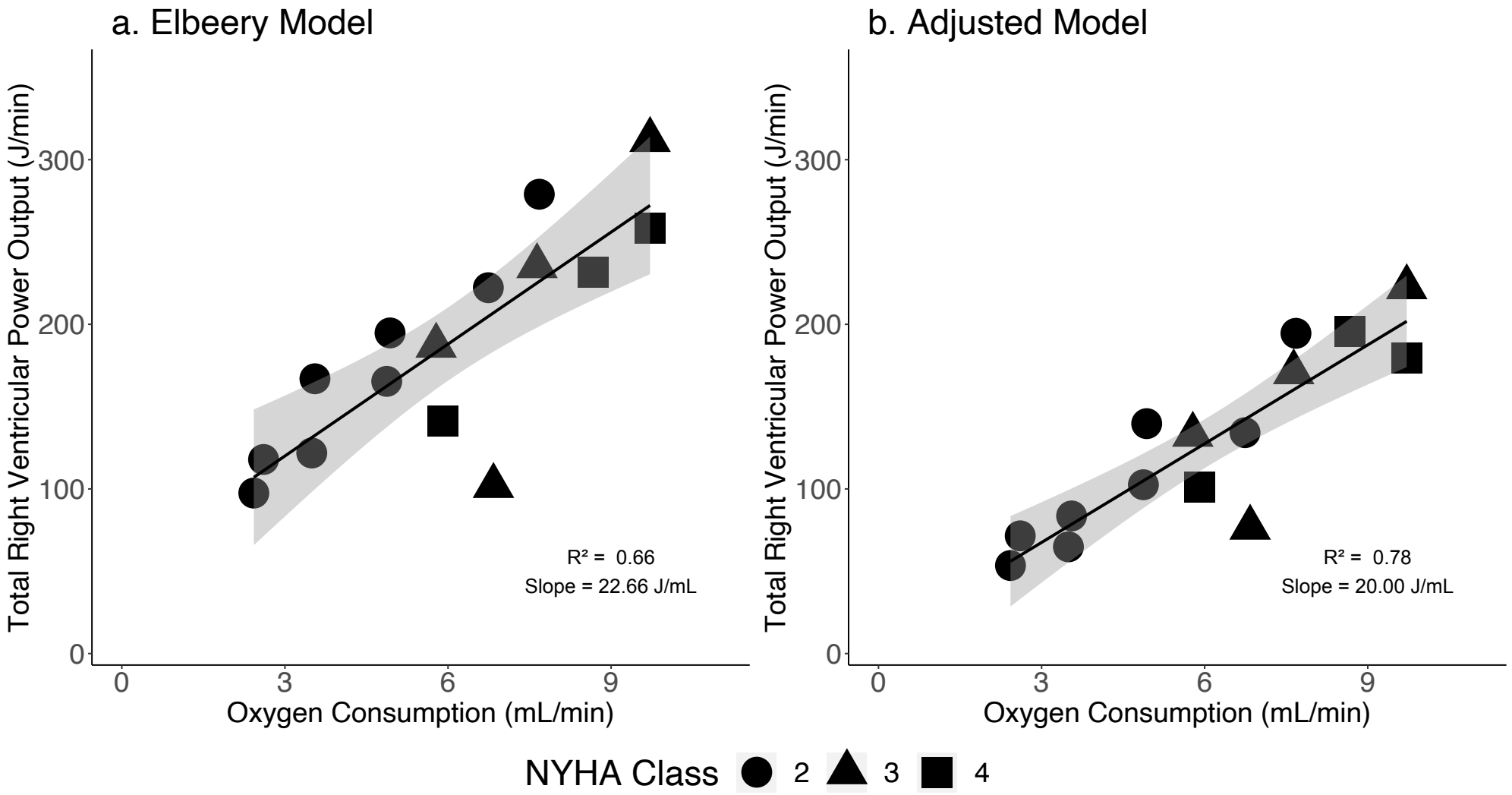


Figure 5.1: Total Right Ventricular Power Output versus Right Myocardial Oxygen Consumption for the Elbeery et al Model (a) and Adjusted Model (b)

As shown in Figure 5.1, the adjusted model had a higher coefficient of determination (*Elbeery et al* model $R^2 = 0.66$ versus adjusted model $R^2 = 0.78$). The slope for the adjusted model (20.0 J/mL O_2) was slightly closer to the theoretical ratio of oxygen to energy output (20.2 J/mL O_2) than the *Elbeery et al* model (22.66 J/mL O_2). The y-intercept for both models, a measurement of power output in the absence of O_2, was not significantly different from zero in either model, but was markedly reduced in the adjusted model (*Elbeery et al* model y-intercept = 52.02, $p = 0.09$ versus adjusted model y-intercept = 7.48, $p = 0.70$). Overall, these estimates account for, on average, 27.7% of total right ventricular power estimated as derived from anaerobic sources in the *Elbeery et al* model versus 6.4% in the adjusted model, respectively.

Figure 5.2 shows the linear relationship between O_2 Consumption and right ventricular total power output for the animal studies using *Elbeery et al* 's model (Figure 5.2a) and the adjusted model (Figure 5.2b).

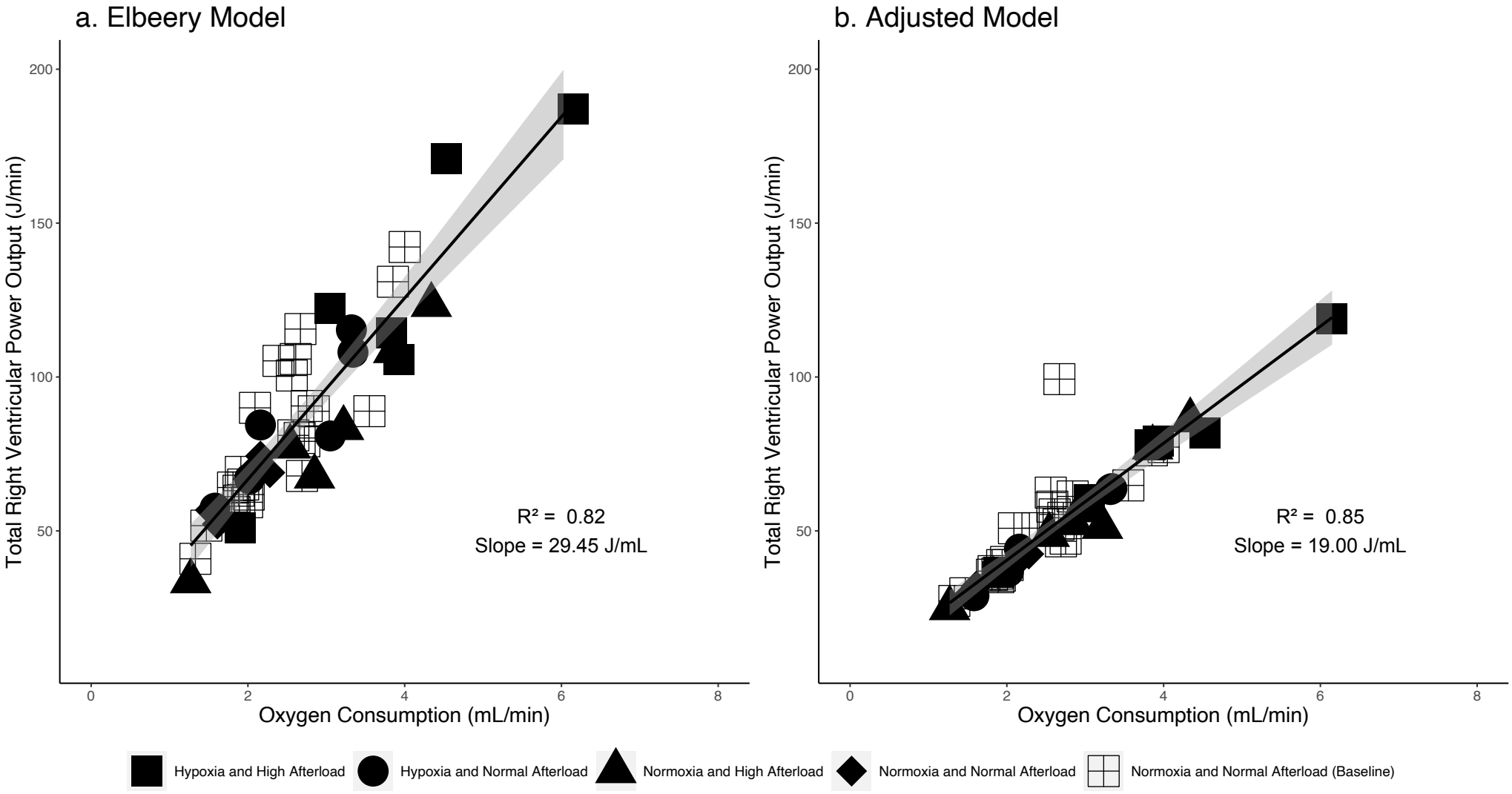
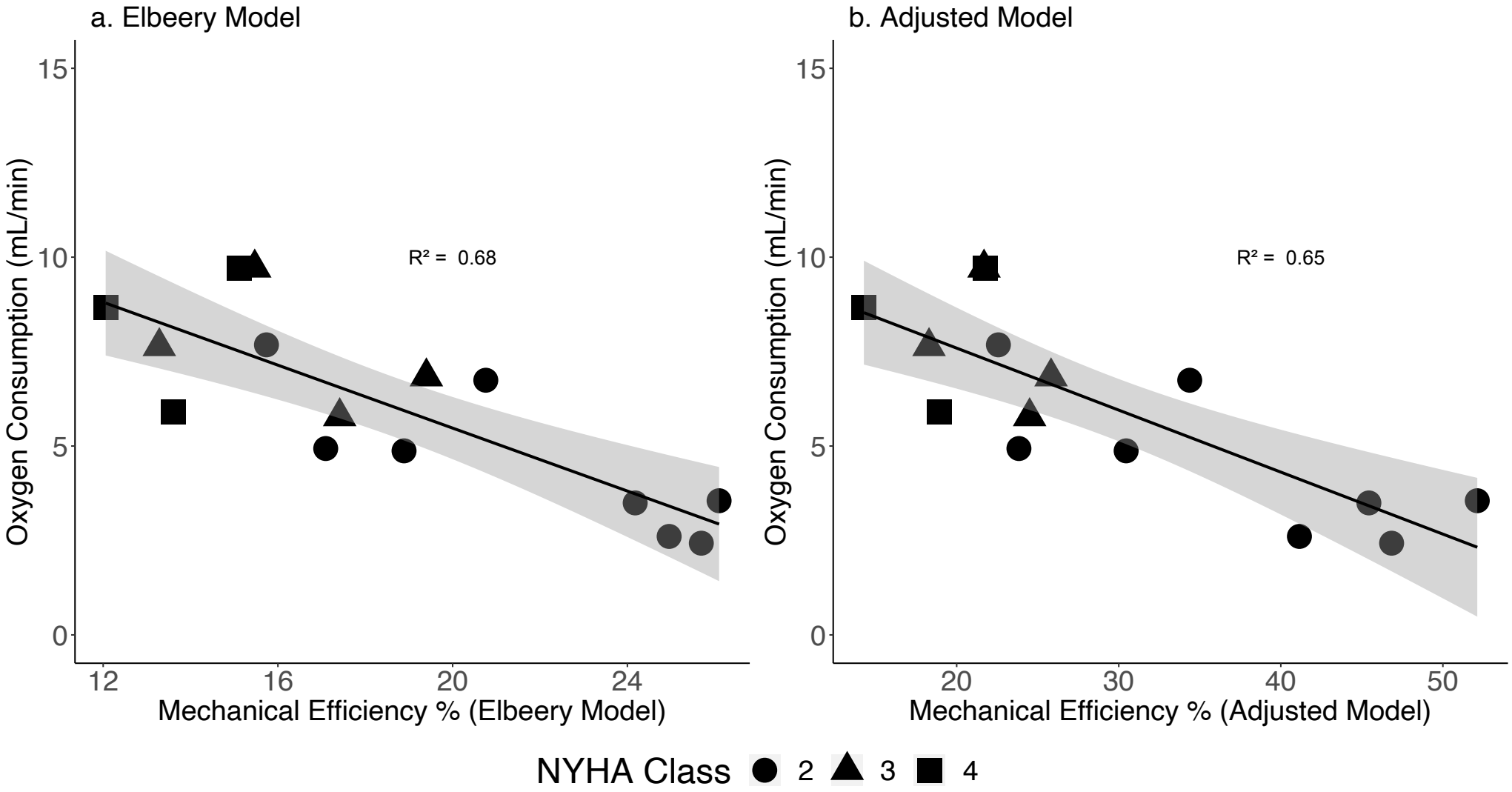

a. Elbeery Model
b. Adjusted Model
Total Right Ventricular Power Output (J/min)
Oxygen Consumption (mL/min)
R² = 0.82
Slope = 29.45 J/mL
R² = 0.85
Slope = 19.00 J/mL
Hypoxia and High Afterload
Hypoxia and Normal Afterload
Normoxia and High Afterload
Normoxia and Normal Afterload
Normoxia and Normal Afterload (Baseline)

Figure 5.2: Total Right Ventricular Power Output vs Right Myocardial Oxygen Consumption for Elbeery et al (a) and Adjusted Model (b)

The adjusted model showed a slightly higher coefficient of determination than the *Elbeery et al* model (*Elbeery et al* model $R^2 = 0.82$ versus adjusted model $R^2 = 0.85$). Again, the slope for the adjusted model (19.00 J/mL O_2) is closer to the theoretical ratio of oxygen to energy output (20.2 J/mL O_2) than the *Elbeery et al* model (29.45 J/mL O_2). Both models did not have a significant y-intercept when applied to the animal study data and again the adjusted model provided a smaller estimate of the y-intercept (*Elbeery et al* model y-intercept = 7.83, p = 0.19 versus adjusted model y-intercept = 2.48, p = 0.47).

Figure 5.3 shows the correlation between RV efficiency (%) and oxygen consumption rate.

Figure 5.3: Right Myocardium Oxygen Consumption vs Right Ventricular Mechanical Efficiency for Elbeery et al Model (a) and



Adjusted Model (b)

The correlation between RV efficiency and oxygen consumption rate was significant for both models ($p < 0.001$). Patients of NYHA Class II were relatively clustered towards having high RV efficiency and low RV oxygen consumption rates, while NYHA Class III/IV were clustered towards having low RV efficiency and high RV oxygen consumption rates.

The average values and 95% confidence intervals for the percent contribution of isovolumic power, useful stroke power, and pulsatile stroke power are provided for NYHA II and III/IV are given in Table 5.3.

Model	Type of Power	NYHA II (n=8)	NYHA III/IV (n=7)
adjusted Model	**ISOVOLUMIC POWER**	52.70±5.51 %	70.68±2.68 %
	PULSATILE STROKE POWER	7.81±1.26 %	7.03±0.81%
	USEFUL STROKE POWER	37.10±3.88 %	20.8±1.48 %
Elbeery Model	**ISOVOLUMIC POWER**	68.00±2.18 %	78.60±1.77 %
	PULSATILE STROKE POWER	9.78±1.27 %	6.22±1.15 %
	USEFUL STROKE POWER	21.68±1.45%	15.19±0.96 %

Table 5.3: Relative Contributions of Different Energy Demands to Total Right Ventricular Power.

Under the *Elbeery et al* model, differences in isovolumic work between NYHA II and III/IV compose the largest difference (78.6 vs 68.0%), with a slight difference as well in pulsatile stroke work. However, in the adjusted model, the differences in pulsatile stroke work are negligible and the differences in isovolumic work are even more substantial (70.6 vs 52.7%). Differences between the estimated useful stroke power is also substantial between NYHA II and NYHA III/IV patients in the adjusted model, but not in the *Elbeery et al* model.

The MANOVA analysis, with Bonferroni correction for multiple comparisons, revealed that the RV efficiency as calculated by the adjusted model and *Elbeery et al* model, and RV oxygen consumption rate all differed significantly between NYHA II versus NYHA III/IV patients ($p = 0.013$, Figure 5.4).

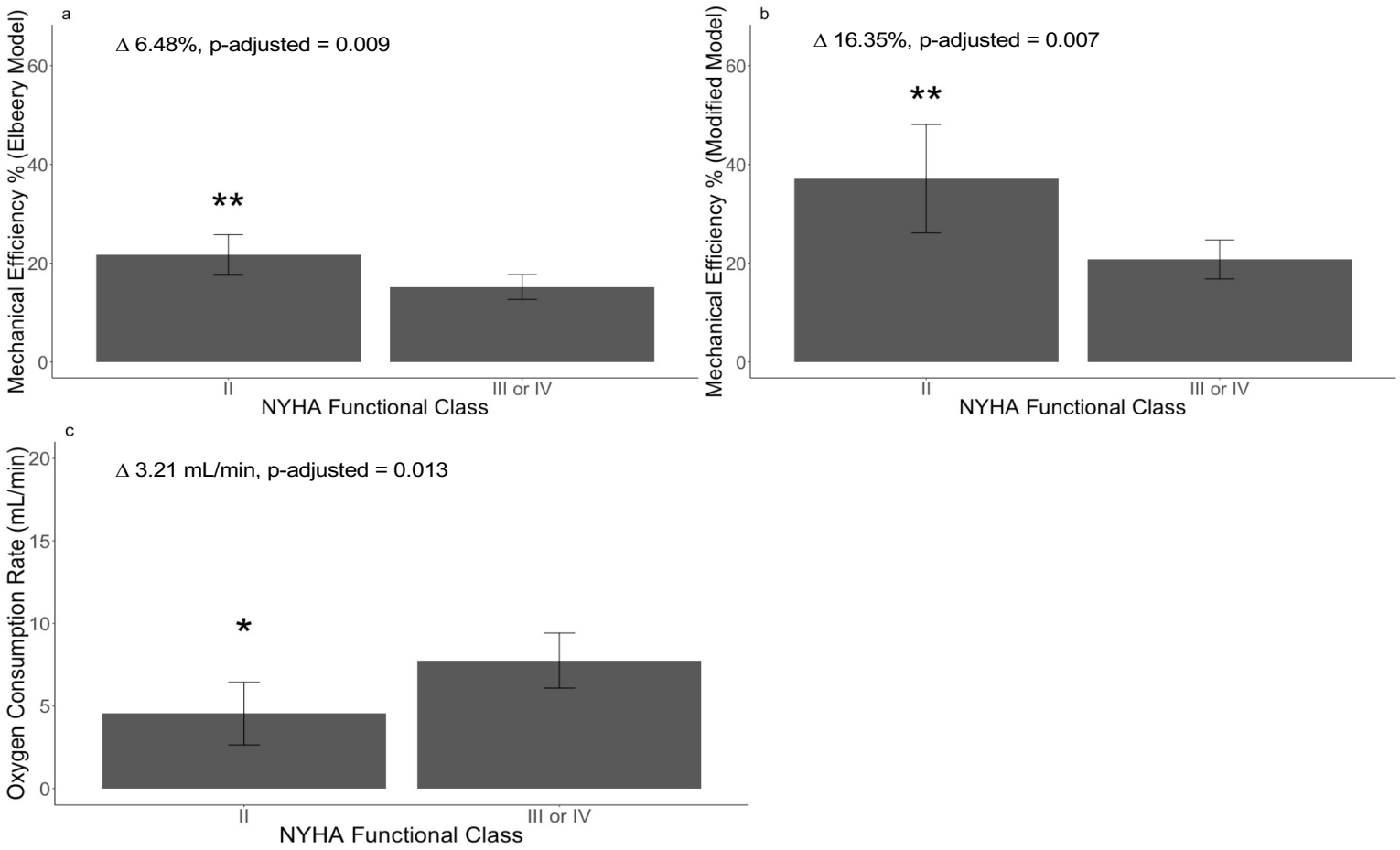


Figure 5.4: Right Ventricular Mechanical Efficiency and Right Myocardial Oxygen Consumption Differences by NYHA Class (II vs III/IV)

Discussion

Two cardiac energetic models are presented that directly couple right ventricular power output and right ventricular myocardial oxygen consumption rate. Both models are novel in their application to the right ventricle under high afterload (i.e. pulmonary arterial hypertension). Adjustments made to the original model published in *Elbeery et al* (1995) improved correlation strength of this coupling when applied to both sheep and human RV data, while also bringing the slope relating RV power and oxygen consumption closer to its theoretical value (20.2 joule per mL O_2). Both models demonstrated that the estimations of total power explain 66-78% of the variance, based on R^2, in oxygen consumption for clinical studies, and 82-85% of the variance in oxygen consumption for animal studies. The first publication on RV energetic inefficiency related to PAH, *Wong et al(122)*, initially cited four other explanations for RV energetic inefficiency, including “tricuspid regurgitation, septal bowing, asynchronous activation, and/or diastolic dysfunction”, but did not identify a large, significant correlation between these effects and overall RV oxygen consumption. Our model demonstrates that, in fact, the majority of right ventricular energetic inefficiency can be explained by hemodynamics and ventricular volumes, while these other explanations might compose a smaller percentage of the variance (approximately 20%). Both models had nonsignificant y-intercept terms, with the adjusted model reducing the value further (down to 6.4%, versus 27.7% in the *Elbeery et al* model), agreeing more closely with the relatively small contribution of anaerobic energy production (< 5%) seen in typical human or sheep myocardial tissue.

Right ventricular efficiency, as shown by the results of the MANOVA, has a clear relationship with disease severity, with more severe right heart failure symptoms corresponding with lower RV efficiency and higher myocardial oxygen consumption. The more energy that is

required to raise the end-diastolic volume to the pulmonary arterial pressure, i.e. right ventricular isovolumic work, the less efficient the RV becomes. The percentage of total power expended towards pulsatile stroke work was roughly equal for patients of low (NYHA Class II) and high disease (NYHA Class III/IV) severity, but the percentage of total power expended towards isovolumic work was much lower for low severity patients versus high severity patients (adjusted model estimates 52% versus 75%, respectively). As raising the end-diastolic volume to the mean PA pressure is a necessary precursor to RV ejection, this initial energy consumption limits the total energy available for stroke work and makes it more difficult for the RV to achieve sufficient stroke volume or cardiac output. Alternatively, using the assumptions made by the adjusted model, if all right ventricular work can be estimated primarily by raising the RV end-diastolic volume to the mean RV ejection pressure, higher RV end-diastolic volume translates to lower possible pressures that can be achieved by the RV. Treatment goals should, therefore, focus on reducing isovolumic power output. Specifically, reduction of right ventricular end-diastolic volume and mean right ventricular ejection pressure should be therapeutic targets for improving RV efficiency and decreasing required oxygen consumption. Further, patients with high isovolumic demands will be less capable of tolerating mild hypoxemia and may be good candidates for oxygen therapy(10)'.

The 2015 European Respiratory Society guidelines on treatment of PAH provide simple risk levels for single hemodynamics such as cardiac index, mean right atrial pressure, and mixed venous oxygen saturation(10). However, the guidelines do not provide a comprehensive assessment of how the relationships between these variables translate to risk or how to combine them into a single quantitative risk score. While there have been multiple attempts to develop risk scores that depend on both hemodynamics and other clinical variables, the most commonly used and often cited scores (REVEAL, French Score, COMPERA) have significant

limitations in their accuracy, as determined by receiver-operator curve analysis(61; 134). Further, all risk scores are based on statistical associations of risk, rather than a deeper physiological understanding of why, for example, a high cardiac index (“low risk”) coupled with a high mean right atrial pressure (“high risk”) should translate to a specific total risk measure.

By contrast, the results of this energetic model show a physiologic justification for combining hemodynamic into comprehensive values that can demonstrate overall RV function (RV efficiency and oxygen consumption). This preliminary study demonstrated a significant association between disease severity and RV oxygen consumption and RV isovolumic work demands. As these values reflect a physiologically-driven combination of crucial hemodynamics, they could potentially provide improvements to risk stratification. By expanding physiological knowledge of the behavior of the right ventricle under the progression of PAH, this model provides insights into the causal mechanisms of increased oxygen extraction and how high isovolumic power reduces energy available for stroke power, thereby reducing cardiac output. Evidence of the importance of improving physiological domain knowledge for development of statistical or machine learning clinical risk models is well-documented(8; 135).

As there is a clear causal mechanism between limited right ventricular oxygen reserve and right heart failure and reduced exercise capacity, it’s extremely likely that these variables would be beneficial to any PAH risk score. Sudden cardiac death remains a frequent cause of death (35-50%) for patients with PAH, which could be explained by right ventricular oxygen reserve finally crossing a threshold, though it may have been at subclinical levels for months or years before that time.(47) Given an adequate patient sample, a Greedy Thick Thinning Bayesian network would be a useful means of studying the causality between exercise capacity (i.e. six minute walk distance), right ventricular oxygen reserve, hemodynamics and mortality.

In the search for novel PAH treatment, right ventricular energetics must be considered. Failure to consider them can have disastrous consequences. For example, efforts to improve RV contractility for PAH patients through certain inotropic drugs such as dobutamine have failed specifically because they increase heart rate while not reducing preload or afterload, thereby reducing oxygen extraction reserve further and metabolically straining cardiomyocytes.(136) In contrast, calcium channel blockers, such as diltiazem, slow cardiac conduction and contractility in vasoreactive patients, reducing RV work demands and providing improved oxygen consumption reserve.(10) Further, diuretics reduce RV end-diastolic volume, allowing for energetically favorable improvements in isovolumic power. An analysis of how pharmaceutical treatments prescribed for PAH may, in isolation or in tandem, reduce RV energetic strain could help explain observed improvements in clinical outcomes that cannot be explained by any single hemodynamic variable alone.

Novel proposed treatments for severe PAH, such as extracorporeal membrane oxygenation (ECMO), also must consider how to optimize RV oxygen delivery and avoid increasing RV work demands(137; 138). Given the number of different attachment modes for ECMO, optimization of the circuit is crucial for providing the greatest therapeutic benefit, especially in situations for the use of ECMO as a bridge-to-lung transplant. Restorative rest for the right ventricle and improving systemic perfusion will improve transplant outcomes and have further potential for bridge-to-recovery(10; 120). This improved energetic model will provide a more accurate quantification of RV offloading than the pressure-volume area models that are currently used(139; 140).

This model further underscores the tenuous circumstance of living with PAH and coronary artery disease. As stated, an estimated 28% of PAH patients also suffer from coronary artery disease.(120) PAH patients are dependent on higher RV oxygen consumption when

afterload is advanced, and in severe PAH, maximal dilation of the right coronary is likely required to achieve the oxygen delivery necessary to maintain sufficient cardiac output. Coronary blockage could prevent patients with CAD from achieving adequate right coronary flow and will result in reduced cardiac output compared to patients with higher coronary flow, even given a similar afterload. For these patients, adequate monitoring and management of blockage is especially crucial. Angiography may be appropriate to explain low cardiac output in situations where neither RV EDV nor RV ejection pressure is exceedingly high. Further, given the greater energetic demands of the RV, these patients may require a lower threshold for percent blockage of the right coronary to become candidates for coronary artery bypass.

Ultimately, for all PAH patients with progression in afterload, the compensatory increase in oxygen consumption will eventually hit a ceiling, as the right coronary artery can only dilate so far and deliver so much oxygen. This is especially true due to limitations in right coronary dilatation and flow caused by increased RV wall tension and increased right atrial pressure. Further increases in mean RV ejection pressure, venous congestion, and decreases in arterial oxygen saturation would thus eventually lead to RV failure. The ability to quantify RV oxygen consumption reserve (difference between maximal RV oxygen consumption and minimal RV oxygen consumption at rest) could provide a means of predicting how close a patient is to entering right heart failure. As oxygen consumption and RV efficiency are related, measurements of peak RV mechanical efficiency during exercise could also provide insight on how well a patient will be able to tolerate physical stress. Future work on this model will, therefore, include investigating if RV oxygen consumption reserve or mechanical efficiency reserve are indicative of survival or a patient's functional capacity (i.e. six-minute walk distance).

Finally, this model demonstrated that anaerobic energy sources for the RV myocardium of PAH patients were negligible, which disagrees somewhat with current thinking on this

topic.(118) There is evidence that a metabolic switch occurs in hypertrophied myocardial tissue where anaerobic energy sources become more heavily favored, but the relative degree to which it contributes to the overall RV energy supply has not been thoroughly studied. The adjusted model determined that this component of RV metabolism is negligible, but the *Elbeery et al* model instead suggests that anaerobic contributions are significant (~27%). At this point, it's unclear if this reflects a significant error in the *Elbeery et al* model, which appears to overestimate total RV power output overall, or if this reflects a physiological phenomenon. In this model, it would be expected that more severe patients would experience higher levels of aerobic glycolysis, translating to a lower total power output than expected given the relative inefficiency of glycolysis versus oxidative phosphorylation. However, as shown in *Wong et al*'s original publication, there was no significant difference in RV myocardial glucose consumption rate between NYHA II and NYHA III patients.(122) This suggests that inefficient oxygen use in NYHA III/IV patients is not primarily due to a greater utilization of aerobic glycolysis. Recent results of treatment of PAH with dichloroacetate intended to reduce mitochondrial decoupling and rate of aerobic glycolysis have demonstrated significant results only in patients with genetic susceptibility.(141) Therefore, patients with a significant metabolic switch likely represent a much smaller percentage of the overall PAH population, while the models discussed here are more widely applicable.

Limitations

Both *Elbeery et al*'s model and the adjusted model have the following limitations:

1) Neither model accounts for stroke work that does not contribute to flow into the pulmonary artery due to tricuspid valve regurgitation. As a common comorbidity of PAH, this additional stroke work may be a key component to gaining a more accurate

representation of total right ventricular power output.(10; 122) Particular to our study, a third

of all patients had moderate or severe valve regurgitation (NYHA II: n = 1, NYHA III: n = 4). This would lead to a small but systematic underestimation of total power output that, while not significant on average (roughly 0.3% of total work), could be more significant on an individual basis (up to 30% of total work). This could potentially be a significant contributor to the high residual for one patient in particular (NYHA Class III), whose O_2 consumption rate overestimated RV total power output in the adjusted model by approximately 45%. However, it is also possible that such outliers are due to isolated measurement error in myocardial oxygen consumption via PET imaging.

2) Neither model accounts explicitly for post-systolic isovolumic contraction, a common phenomenon in PAH that is believed to further contribute to RV energetic inefficiency.(142)

Specifically, this occurs when the RV continues to contract after the pulmonic valve as closed, prolonging the period prior to tricuspid valve opening. As echocardiography data was not available in either study, it's infeasible to more finely delineate between pulmonic valve closure and RV relaxation (i.e. period immediately following RV peak shortening). The model here inferred that RV end-systole occurs during the trough of the first derivative of the RV pressure waveform, and this corresponded strongly with the end of the T-wave in the simultaneously measured ECG. Therefore, the models likely capture the full period of RV contraction and would account for post-systolic isovolumic contraction, but this remains speculative. Regardless, post-systolic isovolumic contraction does not appear to contribute significantly to the error in total RV power output, as RV total power output still explains the majority of the variance in RV oxygen

consumption.

3) Due to the infeasibility of PET imaging for animal studies, two values are used for estimation of oxygen extraction fraction (OEF). These two values were defined in the literature and did not produce significant error in the model, though a more precise model of OEF under different hypoxic conditions would be ideal. However, this ultimately had a negligible effect on the empirical estimation of the efficiency of oxidative phosphorylation (19.0 J/mL in the adjusted model versus 20.2 for the theoretical value).(125)3) Due to the infeasibility of PET imaging for animal studies, two values are used for estimation of oxygen extraction fraction (OEF). These two values were defined in the literature and did not produce significant error in the model, though a more precise model of OEF under different hypoxic conditions would be ideal. However, this ultimately had a negligible effect on the empirical estimation of the efficiency of oxidative phosphorylation (19.0 J/mL in the adjusted model versus 20.2 for the theoretical value).(125)

Conclusion

In conclusion, both models comprehensively describe right ventricular function under high afterload, in a fashion that is physiologically rigorous. We found assumptions made for the second model allowed for greater simplicity in calculation of total RV power output, while maintaining accuracy in measured oxidative phosphorylation efficiency and marginally improving correlation strength between power output and RV oxygen consumption. Both models combine multiple measurements of hemodynamics and right ventricular volumes to create hypothesis-driven biomarkers of disease severity (RV efficiency and RV isovolumic work). In translation to the clinic, RV efficiency could be an important indicator of treatment efficacy, as

both models demonstrate that increases in RV efficiency allow for a decreased requirement for myocardial oxygen consumption at rest. Further studies are required to determine if RV efficiency and/or RV isovolumic work are predictors of clinical outcomes.

Chapter 6 : Conclusion

This dissertation examined the numerous ways in which accurate risk stratification can significantly improve clinical decision support, specifically for pulmonary arterial hypertension (PAH). After providing background on the clinical challenges of both drug efficacy trial design and treatment strategies for PAH in Chapter 1, Chapter 2 examined how current risk stratification tools could be applied to reduce the costs, length, and patient burden of PAH clinical trials. Chapter 3 then examined how cutting-edge risk stratification tools via the recently published Pulmonary Hypertension Outcomes Risk Assessment (PHORA) machine learning model could be employed to examine the risk-benefit tradeoff of multiple oral therapies. After motivating the power of these risk stratification tools, Chapter 4 examined how more objective and quantitative approaches using machine learning enhanced by differential evolution could create a more accurate machine learning tool than PHORA. Finally, Chapter 5 examined a physiological model to enhance clinical understanding of the mechanisms and warning signs of right heart failure via energetic modeling.

6.2 Risk Enrichment for Pulmonary Arterial Hypertension Clinical Trials

- Phase III clinical trials for pulmonary arterial hypertension treatments increasingly employ the use of time to clinical worsening as a primary endpoint, but event-driven trials are lengthy and require large patient enrollments (N > 500) to statistically power. The use of previously published clinical PAH risk assessment

tools as a means of “prognostic enrichment” (which is the practice of enrolling patients more likely to experience an endpoint, i.e. “high risk”) was analyzed.

- Commonly used PAH risk stratification tools (COMPERA, French score, and REVEAL 2.0) were determined to be predictive of clinical worsening events, and therefore can be employed for prognostic enrichment of PAH trials
- REVEAL 2.0 was found to be the most precise in its prediction of clinical worsening, allowing it to significantly reduce required sample sizes in simulated scenarios for two out of three retrospective clinical trials, but lacked accurate inputs for the third trial
- A strategy of enrolling patients who are intermediate to high risk allows for a reduction in required sample sizes and trial length while also keeping screening costs low
- Prognostic enrichment of PAH clinical trials is recommended to reduce time-to-market for lifesaving drugs, reduce time for patients spent on placebo, and decrease drug development costs

6.3 Risk-Benefit Tradeoff of Upfront Combination Therapy for Monotherapy for PAH patients

- Although a single previous clinical trial (AMBITION) demonstrated the potential benefit of treating patients with PAH using oral combination therapy versus monotherapy, it remains unclear if low-risk PAH patients truly benefit from such

aggressive treatment at the cost of experiencing more likely experiencing drug side effects. This hypothesis was examined with a new PAH risk stratification tool, Pulmonary Hypertension Outcomes Risk Assessment (PHORA), the first published and validated machine learning model for PAH risk prediction.

- PHORA was compared to REVEAL [1.0], as a previous publication had suggested that even low risk patients stratified by REVEAL experienced a significant benefit from upfront combination therapy, but REVEAL 1.0 is now considered outdated.
- PHORA was found to be a better predictor of low risk via Net Reclassification Indices, identifying a larger population of low-risk patients when applied retrospectively in the AMBITION clinical trial.
- When separated into low and intermediate-high risk, the AMBITION patient population demonstrated significant treatment heterogeneity, where low risk patients experienced no benefit in delay of clinical worsening and saw a greater likelihood of experiencing an adverse drug effect ($p = 0.056$). Conversely, intermediate, and high-risk patients saw a significant delay in clinical worsening and no significant increased risk of side effects.
- Further, no risk group experienced a benefit in significant delay of mortality on upfront combination therapy versus monotherapy.
- This analysis demonstrates that, when accurately identified, low risk patients may experience a greater propensity for harm than benefit on double combination oral

therapy, and that high-risk patients may benefit more significantly from triple oral therapy or intravenous therapies.

6.3 Improving PAH Risk Stratification with Machine Learning

- A bottom-up approach for developing a novel machine learning model (augmented naïve Bayesian network) for PAH risk assessment was conducted.
- Feature selection for the machine learning model was conducted through a thorough process of 1) meta-analyses, 2) decision tree analysis, 3) differential evolution.
- The final model was found to outperform PHORA as well as all other currently used PAH clinical risk assessment tools (AUC = 0.84).
- The feature selection process revealed several novel biomarkers for PAH mortality risk, including multiple markers of kidney and liver function
- Bayesian networks provide significant promise for improving clinical understanding and knowledge discovery of the systemic consequences of right heart failure caused by PAH.

6.4 Energetic Model of Right Ventricular Failure

- More than 50% of PAH patients die from right heart failure, but predicting when this will occur continues to elude clinicians and researchers. Therefore, an energetic model of the right ventricle, based on the 1995 *Elbeery et al* model, was examined to determine its validity and usefulness in prediction of right heart failure symptoms due to PAH.

- The *Elbeery et al* model was adjusted to increase its clinical relevance to PAH and ease of use.
- Both the *Elbeery et al* model and the adjusted model were found to be highly predictive of RV total power output, as determined by their strong liner relationship with RV myocardial oxygen consumption rate.
- The adjusted model was determined to be more accurate than the *Elbeery et al* model, as its empirical measurement of the efficiency of oxidative phosphorylation.
- Patients with a high degree of right heart failure symptoms (New York Heart Association Functional Class III/IV) were found to have significantly higher right ventricular myocardial oxygen consumption and reduced right ventricular work efficiency.
- Overall, clinical monitoring of right ventricular oxygen reserve and work demands could serve as an early warning sign of right heart failure for patients with PAH.

References

1. Institute of Medicine Committee on Quality of Health Care in A. 2000. In *To Err is Human: Building a Safer Health System*, ed. LT Kohn, JM Corrigan, MS Donaldson. Washington (DC): National Academies Press (US)

2. Rodwin BA, Bilan VP, Merchant NB, Steffens CG, Grimshaw AA, et al. 2020. Rate of Preventable Mortality in Hospitalized Patients: a Systematic Review and Meta-analysis. 35:2099-106
3. Fogel DB. 2018. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. *Contemporary clinical trials communications* 11:156-64
4. Osheroff J, Teich J, Levick D, Saldana L, Velasco F, et al. 2012. *Improving Outcomes with Clinical Decision Support: An Implementer's Guide, Second Edition*.
5. Dinov ID, Petrosyan P, Liu Z, Eggert P, Zamanyan A, et al. 2014. The perfect neuroimaging-genetics-computation storm: collision of petabytes of data, millions of hardware devices and thousands of software tools. 8:311-22
6. Goodson JD, Shahbazi S, Rao K, Song Z. 2020. Differences in the Complexity of Office Visits by Physician Specialty: NAMCS 2013–2016. 35:1715-20
7. Jensen PB, Jensen LJ, Brunak S. 2012. Mining electronic health records: towards better research applications and clinical care. 13:395-405
8. Gopal PSaERaAAaCAaPBaS-HCaPaNDaPEaCEa. 2020. Physiology as a Lingua Franca for Clinical Machine Learning. *Patterns* 1:100017
9. 2020. Mechanistic Modeling and Multiscale Applications for Precision Medicine: Theory and Practice. *Network and Systems Medicine* 3:36-56
10. Galiè N, Humbert M, Vachiery J-L, Gibbs S, Lang I, et al. 2015. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension. <*span class="subtitle">The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS)Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT)</span*> 46:903-75
11. Sommer N, Ghofrani HA, Pak O, Bonnet S, Provencher S, et al. 2021. Current and future treatments of pulmonary arterial hypertension. *British Journal of Pharmacology* 178:6-30
12. Sitbon O, Gomberg-Maitland M, Granton J, Lewis MI, Mathai SC, et al. 2019. Clinical trial design and new therapies for pulmonary arterial hypertension. *The European respiratory journal* 53:1801908
13. Farber HW, Miller DP, Poms AD, Badesch DB, Frost AE, et al. 2015. Five-Year outcomes of patients enrolled in the REVEAL Registry. *Chest* 148:1043-54
14. Galiè N, Olschewski H, Oudiz RJ, Torres F, Frost A, et al. 2008. Ambrisentan for the treatment of pulmonary arterial hypertension: results of the ambrisentan in pulmonary arterial hypertension, randomized, double-blind, placebo-controlled, multicenter, efficacy (ARIES) study 1 and 2. *Circulation* 117:3010-9
15. Rubin LJ, Badesch DB, Barst RJ, Galiè N, Black CM, et al. 2002. Bosentan Therapy for Pulmonary Arterial Hypertension. *New England Journal of Medicine* 346:896-903
16. Pulido T, Adzerikho I, Channick RN, Delcroix M, Galiè N, et al. 2013. Macitentan and Morbidity and Mortality in Pulmonary Arterial Hypertension. *New England Journal of Medicine* 369:809-18

17. Jing ZC, Yu ZX, Shen JY, Wu BX, Xu KF, et al. 2011. Vardenafil in pulmonary arterial hypertension: a randomized, double-blind, placebo-controlled study. *Am J Respir Crit Care Med* 183:1723-9
18. Galiè N, Brundage BH, Ghofrani HA, Oudiz RJ, Simonneau G, et al. 2009. Tadalafil therapy for pulmonary arterial hypertension. *Circulation* 119:2894-903
19. Galiè N, Ghofrani HA, Torbicki A, Barst RJ, Rubin LJ, et al. 2005. Sildenafil Citrate Therapy for Pulmonary Arterial Hypertension. *New England Journal of Medicine* 353:2148-57
20. Ghofrani H-A, Galiè N, Grimminger F, Grünig E, Humbert M, et al. 2013. Riociguat for the Treatment of Pulmonary Arterial Hypertension. *New England Journal of Medicine* 369:330-40
21. Ghofrani HA, Galiè N, Grimminger F, Grünig E, Humbert M, et al. 2013. Riociguat for the treatment of pulmonary arterial hypertension. *N Engl J Med* 369:330-40
22. Galiè N, Müller K, Scalise A-V, Grünig E. 2015. PATENT PLUS: a blinded, randomised and extension study of riociguat plus sildenafil in pulmonary arterial hypertension. *European Respiratory Journal* 45:1314-22
23. Rubin LJ, Mendoza J, Hood M, McGoon M, Barst R, et al. 1990. Treatment of primary pulmonary hypertension with continuous intravenous prostacyclin (epoprostenol). Results of a randomized trial. *Ann Intern Med* 112:485-91
24. Olschewski H, Simonneau G, Galiè N, Higenbottam T, Naeije R, et al. 2002. Inhaled iloprost for severe pulmonary hypertension. *N Engl J Med* 347:322-9
25. Sitbon O, Vonk Noordegraaf A. 2017. Epoprostenol and pulmonary arterial hypertension: 20 years of clinical experience. *European Respiratory Review* 26:160055
26. Higenbottam T, Butt AY, McMahon A, Westerbeck R, Sharples L. 1998. Long term intravenous prostaglandin (epoprostenol or iloprost) for treatment of severe pulmonary hypertension. *Heart* 80:151-5
27. Higenbottam TW, Butt AY, Dinh-Xaun AT, Takao M, Cremona G, Akamine S. 1998. Treatment of pulmonary hypertension with the continuous infusion of a prostacyclin analogue, iloprost. *Heart* 79:175-9
28. Provencher S, Sitbon O. 2009. Intravenous iloprost for pulmonary arterial hypertension: still waiting for evidence. *European Respiratory Journal* 34:7-9
29. Chen YF, Jowett S, Barton P, Malottki K, Hyde C, et al. 2009. Clinical and cost-effectiveness of epoprostenol, iloprost, bosentan, sitaxentan and sildenafil for pulmonary arterial hypertension within their licensed indications: a systematic review and economic evaluation. *Health Technol Assess* 13:1-320
30. Administration USFaD. 2014. The Voice of the Patient: A series of reports from the U.S. Food and Drug Administration's (FDA's)

Patient-Focused Drug Development Initiative, Pulmonary Arterial Hypertension. Online: United States Government

31. Ogawa A, Satoh T, Tamura Y, Fukuda K, Matsubara H. 2017. Survival of Japanese Patients With Idiopathic/Heritable Pulmonary Arterial Hypertension. *American Journal of Cardiology* 119:1479-84
32. Simonneau G, Barst RJ, Galie N, Naeije R, Rich S, et al. 2002. Continuous subcutaneous infusion of treprostinil, a prostacyclin analogue, in patients with pulmonary arterial hypertension: a double-blind, randomized, placebo-controlled trial. *Am J Respir Crit Care Med* 165:800-4

33. Galiè N, Humbert M, Vachiéry JL, Vizza CD, Kneussl M, et al. 2002. Effects of beraprost sodium, an oral prostacyclin analogue, in patients with pulmonary arterial hypertension: a randomized, double-blind, placebo-controlled trial. *J Am Coll Cardiol* 39:1496-502
34. Tapson VF, Jing ZC, Xu KF, Pan L, Feldman J, et al. 2013. Oral treprostinil for the treatment of pulmonary arterial hypertension in patients receiving background endothelin receptor antagonist and phosphodiesterase type 5 inhibitor therapy (the FREEDOM-C2 study): a randomized controlled trial. *Chest* 144:952-8
35. McLaughlin VV, Oudiz RJ, Frost A, Tapson VF, Murali S, et al. 2006. Randomized study of adding inhaled iloprost to existing bosentan in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 174:1257-63
36. McLaughlin VV, Benza RL, Rubin LJ, Channick RN, Voswinckel R, et al. 2010. Addition of inhaled treprostinil to oral therapy for pulmonary arterial hypertension: a randomized controlled clinical trial. *J Am Coll Cardiol* 55:1915-22
37. Sitbon O, Channick R, Chin KM, Frey A, Gaine S, et al. 2015. Selexipag for the Treatment of Pulmonary Arterial Hypertension. *New England Journal of Medicine* 373:2522-33
38. Humbert M, Barst RJ, Robbins IM, Channick RN, Galiè N, et al. 2004. Combination of bosentan with epoprostenol in pulmonary arterial hypertension: BREATHE-2. *European Respiratory Journal* 24:353-9
39. Kemp K, Savale L, O'Callaghan DS, Jaïs X, Montani D, et al. 2012. Usefulness of first-line combination therapy with epoprostenol and bosentan in pulmonary arterial hypertension: an observational study. *J Heart Lung Transplant* 31:150-8
40. McLaughlin V, Channick RN, Ghofrani H-A, Lemarié J-C, Naeije R, et al. 2015. Bosentan added to sildenafil therapy in patients with pulmonary arterial hypertension. *European Respiratory Journal* 46:405-13
41. Simonneau G, Rubin LJ, Galiè N, Barst RJ, Fleming TR, et al. 2008. Addition of sildenafil to long-term intravenous epoprostenol therapy in patients with pulmonary arterial hypertension: a randomized trial. *Ann Intern Med* 149:521-30
42. Galiè N, Barberà JA, Frost AE, Ghofrani H-A, Hoeper MM, et al. 2015. Initial Use of Ambrisentan plus Tadalafil in Pulmonary Arterial Hypertension. *New England Journal of Medicine* 373:834-44
43. Sitbon O, Jaïs X, Savale L, Cottin V, Bergot E, et al. 2014. Upfront triple combination therapy in pulmonary arterial hypertension: a pilot study. *Eur Respir J* 43:1691-7
44. Sitbon O, Humbert M, Jaïs X, Ioos V, Hamid AM, et al. 2005. Long-Term Response to Calcium Channel Blockers in Idiopathic Pulmonary Arterial Hypertension. *Circulation* 111:3105-11
45. Sitbon O, Brenot F, Denjean A, Bergeron A, Parent F, et al. 1995. Inhaled nitric oxide as a screening vasodilator agent in primary pulmonary hypertension. A dose-response study and comparison with prostacyclin. *American Journal of Respiratory and Critical Care Medicine* 151:384-9
46. Besinque GM, Lickert CA, Pruett JA. 2019. The myth of the stable pulmonary arterial hypertension patient. *Am J Manag Care* 25:S47-s52
47. Demerouti EA, Manginas AN, Athanassopoulos GD, Karatasakis GT. 2013. Complications Leading to Sudden Cardiac Death in Pulmonary Arterial Hypertension. *Respiratory Care* 58:1246-54
48. Tonelli AR, Arelli V, Minai OA, Newman J, Bair N, et al. 2013. Causes and circumstances of death in pulmonary arterial hypertension. *Am J Respir Crit Care Med* 188:365-9

49. Bustamante-Labarta M, Perrone S, De La Fuente RL, Stutzbach P, De La Hoz RP, et al. 2002. Right atrial size and tricuspid regurgitation severity predict mortality or transplantation in primary pulmonary hypertension. *J Am Soc Echocardiogr* 15:1160-4
50. Sahay S, Tonelli AR. 2013. Pericardial effusion in pulmonary arterial hypertension. *Pulmonary circulation* 3:467-77
51. Khirfan G, Almoushref A, Naal T, Abuhalimeh B, Dweik RA, et al. 2020. Mixed Venous Oxygen Saturation Is a Better Prognosticator Than Cardiac Index in Pulmonary Arterial Hypertension. *Chest* 158:2546-55
52. Mueller T, Gegenhuber A, Poelz W, Haltmayer M. 2003. Comparison of the Biomedica NT-proBNP enzyme immunoassay and the Roche NT-proBNP chemiluminescence immunoassay: implications for the prediction of symptomatic and asymptomatic structural heart disease. *Clin Chem* 49:976-9
53. Schwaiblmair M, Faul C, von Scheidt W, Berghaus TM. 2012. Ventilatory efficiency testing as prognostic value in patients with pulmonary hypertension. *BMC Pulm Med* 12:23
54. Xanthouli P, Koegler M, Marra AM, Benjamin N, Fischer L, et al. 2020. Risk stratification and prognostic factors in patients with pulmonary arterial hypertension and comorbidities a cross-sectional cohort study with survival follow-up. 21:127
55. Dolgin M, New York Heart A, Criteria C. 1994. *Nomenclature and criteria for diagnosis of diseases of the heart and great vessels*. Boston: Little, Brown
56. Taichman DB, McGoon MD, Harhay MO, Archer-Chicko C, Sager JS, et al. 2009. Wide variation in clinicians' assessment of New York Heart Association/World Health Organization functional class in patients with pulmonary arterial hypertension. *Mayo Clin Proc* 84:586-92
57. Raphael C, Briscoe C, Davies J, Ian Whinnett Z, Manisty C, et al. 2007. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. *Heart* 93:476-82
58. Frost AE, Hoeper MM, Barberá JA, Vachiery JL, Blair C, et al. 2018. Risk-stratified outcomes with initial combination therapy in pulmonary arterial hypertension: Application of the REVEAL risk score. *J Heart Lung Transplant* 37:1410-7
59. Hoeper MM, Kramer T, Pan Z, Eichstaedt CA, Spiesshoefer J, et al. 2017. Mortality in pulmonary arterial hypertension: prediction by the 2015 European pulmonary hypertension guidelines risk stratification model. *European Respiratory Journal* 50:1700740
60. Boucly A, Weatherald J, Savale L, Jaïs X, Cottin V, et al. 2017. Risk assessment, prognosis and guideline implementation in pulmonary arterial hypertension. *European Respiratory Journal* 50:1700889
61. Benza RL, Gomberg-Maitland M, Elliott CG, Farber HW, Foreman AJ, et al. 2019. Predicting Survival in Patients With Pulmonary Arterial Hypertension: The REVEAL Risk Score Calculator 2.0 and Comparison With ESC/ERS-Based Risk Assessment Strategies. *CHEST* 156:323-37
62. Benza RL, Miller DP, Gomberg-Maitland M, Frantz RP, Foreman AJ, et al. 2010. Predicting survival in pulmonary arterial hypertension: insights from the Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management (REVEAL). *Circulation* 122:164-72
63. Lewis RA, Johns CS, Cogliano M, Capener D, Tubman E, et al. 2020. Identification of Cardiac Magnetic Resonance Imaging Thresholds for Risk Stratification in Pulmonary Arterial Hypertension. *Am J Respir Crit Care Med* 201:458-68

64. Anderson JJ, Lau EM, Lavender M, Benza R, Celermajer DS, et al. 2020. Retrospective Validation of the REVEAL 2.0 Risk Score With the Australian and New Zealand Pulmonary Hypertension Registry Cohort. *Chest* 157:162-72
65. Kanwar MK, Gomberg-Maitland M, Hoeper M, Pausch C, Pittow D, et al. 2020. Risk stratification in pulmonary arterial hypertension using Bayesian analysis. *European Respiratory Journal*:2000008
66. Raina A, Humbert M. 2016. Risk assessment in pulmonary arterial hypertension. *European Respiratory Review* 25:390-8
67. Administration USFaD. 2019. Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products: Guidance for Industry. ed. CfDEa Research. Online: United States Government
68. Clark TG, Bradburn MJ, Love SB, Altman DG. 2003. Survival analysis part I: basic concepts and first analyses. *British journal of cancer* 89:232-8
69. Kumar D, Klefsjö B. 1994. Proportional hazards model: a review. 44:177-88
70. Austin PC. 2018. Statistical power to detect violation of the proportional hazards assumption when using the Cox regression model. *Journal of Statistical Computation and Simulation* 88:533-52
71. Schmoor C, Sauerbrei W, Schumacher M. 2000. Sample size considerations for the evaluation of prognostic factors in survival analysis. *Stat Med* 19:441-52
72. Friedman N, Geiger D, Goldszmidt M. 1997. Bayesian Network Classifiers. 29:131-63
73. Kang H. 2013. The prevention and handling of the missing data. *Korean J Anesthesiol* 64:402-6
74. Cheng JaBDaLW. 1998. An Algorithm for Bayesian Belief Network Construction from Data. *Proceedings of AI & STAT'97*
75. Pearl J. 1988. Chapter 3 - MARKOV AND BAYESIAN NETWORKS: Two Graphical Representations of Probabilistic Knowledge.77-141. San Francisco (CA): Morgan Kaufmann. Number of 77-141 pp.
76. Spirtes PL, Glymour C, Scheines R. Causation, prediction, and search, *1993*:
77. Storn R, Price K. 1997. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. 11:341-59
78. McKay MD, Beckman RJ, Conover WJ. 1979. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics* 21:239-45
79. Hancer E, Xue B, Zhang M. 2018. Differential evolution for filter feature selection based on information theory and feature ranking. 140:103-19
80. Hassan S, Hemeida AM, Alkhalaf S, Mohamed A-A, Senjyu T. 2020. Multi-variant differential evolution algorithm for feature selection. 10:17261
81. Ratanamahatana CA, Gunopulos D. 2002. Scaling up the naive Bayesian classifier: Using decision trees for feature selection.
82. Breiman L, Friedman JH, Olshen RA, Stone CJ. 2017. *Classification and regression trees*. Routledge
83. Zhou Y, McArdle JJ. 2015. Rationale and Applications of Survival Tree and Survival Ensemble Methods. 80:811-33
84. Mingers J. 1989. An empirical comparison of selection measures for decision-tree induction. 3:319-42
85. Khalafbeigui F, Suga H, Sagawa K. 1979. Left ventricular systolic pressure-volume area correlates with oxygen consumption. *American Journal of Physiology-Heart and Circulatory Physiology* 237:H566-H9

86. Klotz S, Hay I, Dickstein ML, Yi GH, Wang J, et al. 2006. Single-beat estimation of end-diastolic pressure-volume relationship: a novel method with potential for noninvasive application. *Am J Physiol Heart Circ Physiol* 291:H403-12
87. Elbeery JR, Lucke JC, Feneley MP, Maier GW, Owen CH, et al. 1995. Mechanical determinants of myocardial oxygen consumption in conscious dogs. *Am J Physiol* 269:H609-20
88. Loiselle DS, Taberner AJ, Tran K, Han J-C. 2021. Thermodynamic inconsistency disproves the Suga-Sagawa theory of cardiac energetics. 164:81-91
89. Ryan JJ, Rich JD, Maron BA. 2015. Building the case for novel clinical trials in pulmonary arterial hypertension. *Circulation. Cardiovascular quality and outcomes* 8:114-23
90. Kanwar MK, Thenappan T, Vachiéry J-L. 2016. Update in treatment options in pulmonary hypertension. *The Journal of Heart and Lung Transplantation* 35:695-703
91. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, et al. 2009. A new equation to estimate glomerular filtration rate. *Ann Intern Med* 150:604-12
92. Zou KH, O'Malley AJ, Mauri L. 2007. Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation* 115:654-7
93. Therneau T, Atkinson E. 1997. An introduction to Recursive Partitioning Using the RPART Routines. 61
94. Qiu W, Chavarro J, Lazarus R, Rosner B, Ma J. 2018. powerSurvEpi: Power and Sample Size Calculation for Survival Analysis of Epidemiological Studies.
95. Freedman LS. 1982. Tables of the number of patients required in clinical trials using the logrank test. *Stat Med* 1:121-9
96. Scientific B. 2019. Procedural payment guide: hospital inpatient and outpatient, ASC and physician reimbursement information. Online
97. Health CAfDaTi. 2017. Pharmacoeconomic review report: selexipeg (Uptravi). In *Online*
98. Benza RL, Kanwar MK, Raina A, Scott JV, Zhao CL, et al. 2021. Development and Validation of an Abridged Version of the REVEAL 2.0 Risk Score Calculator, REVEAL Lite 2, for Use in Patients With Pulmonary Arterial Hypertension. *Chest* 159:337-46
99. Yang YT, Chen B, Bennett CL. 2018. Offshore Pharmaceutical Trials: Evidence, Economics, and Ethics. *Mayo Clinic proceedings. Innovations, quality & outcomes* 2:226-8
100. Moore TJ, Zhang H, Anderson G, Alexander GC. 2018. Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016. *JAMA Intern Med* 178:1451-7
101. Administration USFaD. 2019. Enrichment Strategies for
Clinical Trials to Support
Determination of
Effectiveness of Human Drugs
and Biological Products: Guidance for Industry. ed. CfDEa Research. Online: United States Government
102. Kerr KF, Wang Z, Janes H, McClelland RL, Psaty BM, Pepe MS. 2014. Net reclassification indices for evaluating risk prediction instruments: a critical review. *Epidemiology (Cambridge, Mass.)* 25:114-21
103. Rosenkranz S, Howard LS, Gomberg-Maitland M, Hoeper MM. 2020. Systemic Consequences of Pulmonary Hypertension and Right-Sided Heart Failure. *Circulation* 141:678-93

104. George MP, Champion HC, Pilewski JM. 2011. Lung transplantation for pulmonary hypertension. *Pulmonary circulation* 1:182-91
105. Marshall A, Altman DG, Holder RL. 2010. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. 10:112
106. Ji Z, Xia Q, Meng G. A Review of Parameter Learning Methods in Bayesian Network, *Cham*, *2015*:3-12: Springer International Publishing
107. Watson DS, Krutzinna J, Bruce IN, Griffiths CE, McInnes IB, et al. 2019. Clinical applications of machine learning algorithms: beyond the black box. *BMJ* 364:l886
108. Oudiz RJ, Galiè N, Olschewski H, Torres F, Frost A, et al. 2009. Long-term ambrisentan therapy for the treatment of pulmonary arterial hypertension. *J Am Coll Cardiol* 54:1971-81
109. White RJ, Jerjes-Sanchez C, Bohns Meyer GM, Pulido T, Sepulveda P, et al. 2020. Combination Therapy with Oral Treprostinil for Pulmonary Arterial Hypertension. A Double-Blind Placebo-controlled Clinical Trial. *Am J Respir Crit Care Med* 201:707-17
110. Rubin LJ, Galiè N, Grimminger F, Grünig E, Humbert M, et al. 2015. Riociguat for the treatment of pulmonary arterial hypertension: a long-term extension study (PATENT-2). *Eur Respir J* 45:1303-13
111. Riley RD, Simmonds MC, Look MP. 2007. Evidence synthesis combining individual patient data and aggregate data: a systematic review identified current practice and possible methods. 60:431.e1-.e12
112. Ferreira DC, Vázquez FI, Vormayr G, Bachl M, Zseby T. A meta-analysis approach for feature selection in network traffic research. *Proc. Proceedings of the Reproducibility Workshop*, *2017*:17-20:
113. Ghofrani HA, D'Armini AM, Grimminger F, Hoeper MM, Jansa P, et al. 2013. Riociguat for the treatment of chronic thromboembolic pulmonary hypertension. *N Engl J Med* 369:319-29
114. Therapeutics U. 2021. An Open-Label Extension Trial of UT-15C Sustained-release (SR) in Subjects With Pulmonary Arterial Hypertension (FREEDOM-EXT). ed. USDoHaH Services. ClinicalTrials.gov: U.S. National Library of Medicine
115. Ferreira B, Silva RG, Pereira V. Feature selection using non-binary decision trees applied to condition monitoring. *Proc. 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, *2017*:1-7:
116. Mingers J. 1989. An Empirical Comparison of Pruning Methods for Decision Tree Induction. 4:227-43
117. Pulmonary Hypertension Association I. Pulmonary Hypertension Association Registry (PHAR). ed. USDoHaH Services. ClinicalTrials.gov: U.S. National Library of Medicine
118. Ryan JJ, Archer SL. 2014. The right ventricle in pulmonary arterial hypertension: disorders of metabolism, angiogenesis and adrenergic signaling in right ventricular failure. *Circulation research* 115:176-88
119. van Wolferen SA, Marcus JT, Westerhof N, Spreeuwenberg MD, Marques KM, et al. 2008. Right coronary artery flow impairment in patients with pulmonary hypertension. *Eur Heart J* 29:120-7
120. Shimony A, Eisenberg MJ, Rudski LG, Schlesinger R, Afilalo J, et al. 2011. Prevalence and impact of coronary artery disease in patients with pulmonary arterial hypertension. *Am J Cardiol* 108:460-4
121. Vogel-Claussen J, Skrok J, Shehata ML, Singh S, Sibley CT, et al. 2011. Right and left ventricular myocardial perfusion reserves correlate with right ventricular function and

pulmonary hemodynamics in patients with pulmonary arterial hypertension. *Radiology* 258:119-27
122. Wong YY, Ruiter G, Lubberink M, Raijmakers PG, Knaapen P, et al. 2011. Right Ventricular Failure in Idiopathic Pulmonary Arterial Hypertension Is Associated With Inefficient Myocardial Oxygen Utilization. *Circulation: Heart Failure* 4:700-6
123. Wong YY, Westerhof N, Ruiter G, Lubberink M, Raijmakers P, et al. 2011. Systolic pulmonary artery pressure and heart rate are main determinants of oxygen consumption in the right ventricular myocardium of patients with idiopathic pulmonary arterial hypertension. *Eur J Heart Fail* 13:1290-5
124. Suga H, Hayashi T, Shirahata M. 1981. Ventricular systolic pressure-volume area as predictor of cardiac oxygen consumption. *Am J Physiol* 240:H39-44
125. Martínez MS, García A, Luzardo E, Chávez-Castillo M, Olivar LC, et al. 2017. Energetic metabolism in cardiomyocytes: molecular basis of heart ischemia and arrhythmogenesis. *Vessel Plus* 1:130-41
126. Crystal GJ, Pagel PS. 2018. Right Ventricular Perfusion: Physiology and Clinical Implications. *Anesthesiology* 128:202-18
127. Likitha L, Gudepu P, Lakshman M. 2018. Gross morphology and morphometry of coronary arteries in sheep (Ovis aries).
128. Cotton BA, McElroy LA. 2015. Oxygen-Carrying Capacity. In *Encyclopedia of Trauma Care*, ed. PJ Papadakos, ML Gestring:1150-. Berlin, Heidelberg: Springer Berlin Heidelberg. Number of 1150- pp.
129. Hart BJ, Bian X, Gwirtz PA, Setty S, Downey HF. 2001. Right ventricular oxygen supply/demand balance in exercising dogs. *Am J Physiol Heart Circ Physiol* 281:H823-30
130. Saito D, Tani H, Kusachi S, Uchida S, Ohbayashi N, et al. 1991. Oxygen metabolism of the hypertrophic right ventricle in open chest dogs. *Cardiovasc Res* 25:731-9
131. Lubberink M, Wong YY, Raijmakers PGHM, Schuit RC, Luurtsema G, et al. 2011. Myocardial Oxygen Extraction Fraction Measured Using Bolus Inhalation of ¹⁵O-Oxygen Gas and Dynamic PET. *Journal of Nuclear Medicine* 52:60-6
132. van Grondelle A. 1982. *Analysis of pulmonary arterial blood flow dynamics with special reference to congenital heart disease*. Northwestern University
133. Saouti N, Westerhof N, Helderman F, Marcus JT, Boonstra A, et al. 2010. Right ventricular oscillatory power is a constant fraction of total power irrespective of pulmonary artery pressure. *Am J Respir Crit Care Med* 182:1315-20
134. Galiè N, Channick RN, Frantz RP, Grünig E, Jing ZC, et al. 2019. Risk stratification and medical therapy of pulmonary arterial hypertension. *European Respiratory Journal* 53:1801889
135. Roe KD, Jawa V, Zhang X, Chute CG, Epstein JA, et al. 2020. Feature engineering with clinical expert knowledge: A case study assessment of machine learning model complexity and performance. *PLoS One* 15:e0231300
136. Groeneveldt JA, de Man FS, Westerhof BE. 2019. The right treatment for the right ventricle. *Current opinion in pulmonary medicine* 25:410-7
137. Machuca TN, Perrot Md. 2015. Mechanical Support for the Failing Right Ventricle in Patients With Precapillary Pulmonary Hypertension. *Circulation* 132:526-36
138. Rosenzweig EB, Brodie D, Abrams DC, Agerstrand CL, Bacchetta M. 2014. Extracorporeal membrane oxygenation as a novel bridging strategy for acute right heart failure in group 1 pulmonary arterial hypertension. *Asaio j* 60:129-33

139. Perlman CE, Mockros LF. 2007. Hemodynamic Consequences of Thoracic Artificial Lung Attachment Configuration: A Computational Model. *ASAIO Journal* 53:50-64
140. Boschetti F, Perlman CE, Cook KE, Mockros LF. 2000. Hemodynamic Effects of Attachment Modes and Device Design of a Thoracic Artificial Lung. *ASAIO Journal* 46:42-8
141. Michelakis ED, Gurtu V, Webster L, Barnes G, Watson G, et al. 2017. Inhibition of pyruvate dehydrogenase kinase improves pulmonary arterial hypertension in genetically susceptible patients. *Sci Transl Med* 9
142. Mauritz GJ, Marcus JT, Westerhof N, Postmus PE, Vonk-Noordegraaf A. 2011. Prolonged right ventricular post-systolic isovolumic period in pulmonary arterial hypertension is not a reflection of diastolic dysfunction. *Heart* 97:473-8

Appendix

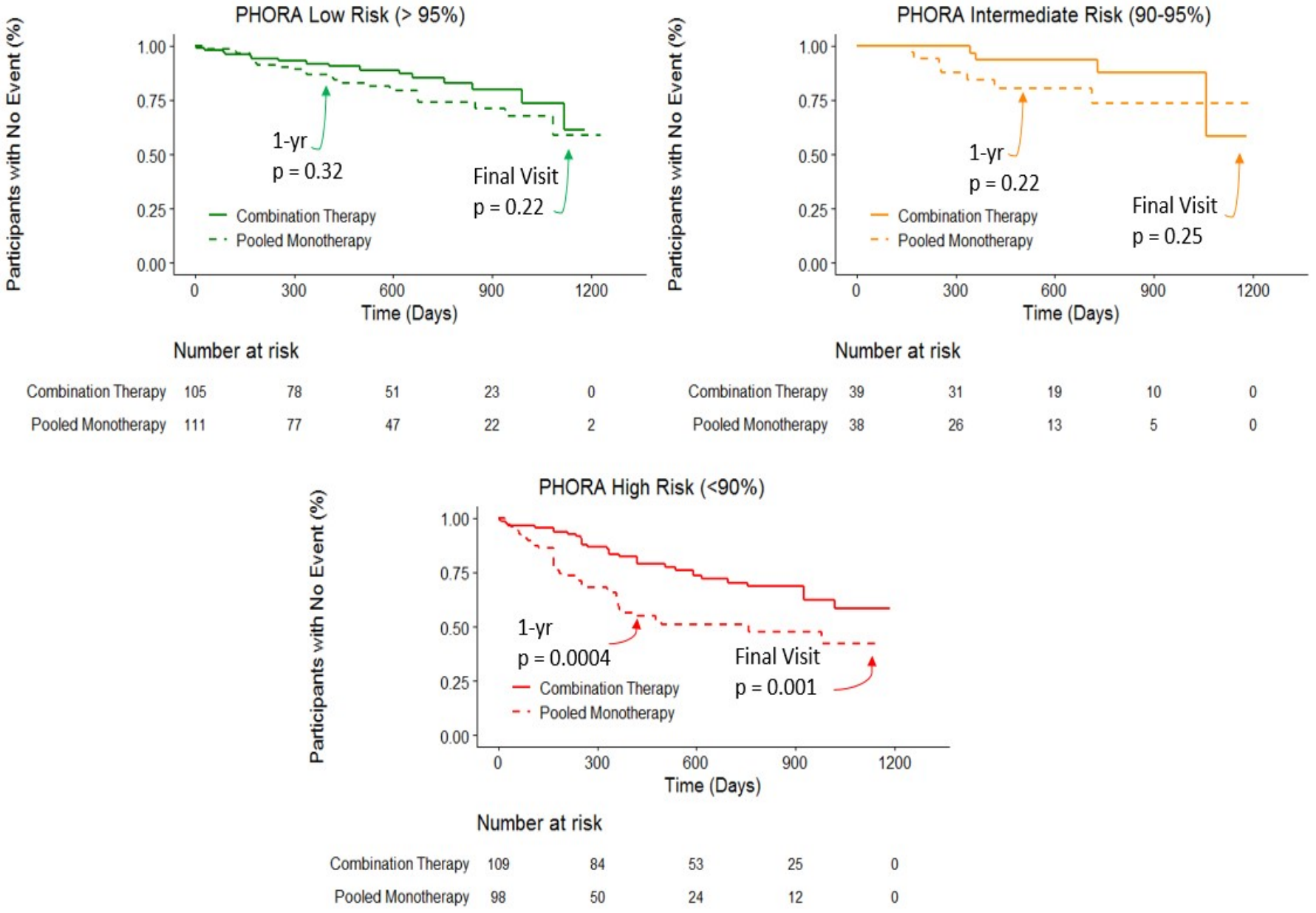


Figure A- 1. Time to Clinical Failure for Combination Therapy versus Monotherapy for PAH patients stratified by PHORA.

Python Code for Feature Preprocessing and Differential Evolution

[1]:
```
from sklearn.model_selection import cross_val_score
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import plot_tree
from sklearn.model_selection import StratifiedKFold
from sklearn.model_selection import GridSearchCV
from scipy.stats import chi2_contingency
import numpy as np
import pickle
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.metrics import roc_auc_score
from pomegranate import *
import miceforest
import copy
from itertools import combinations
%matplotlib inline
from scipy.stats import kendalltau
```

[10]:
```
pd.crosstab(all_data["trial"],all_data["RACE"])
```

[10]:

trial	RACE	ASIAN	BLACK OR AFRICAN AMERICAN	HISPANIC	MIXED	OTHER	WHITE
AMBITION		10	31	0	3	3	395
ARIES		3	5	17	0	1	87
FREEDOM		256	14	0	0	3	311
GRIPHON		232	25	102	0	16	685
PATENT		89	2	5	0	0	141
SERAPHIN		203	19	107	0	2	389

[11]:
```
#Etiology of PAH
pd.crosstab(all_data["trial"],all_data["DIAG_CD"])
```

[11]:

trial	DIAG_CD	APAH-CHD	APAH-CVCTD	APAH-HIV	APAH-TOX	FPAH	IPAH	PORTAL
AMBITION		6	167	8	13	11	238	0
ARIES		0	38	5	1	0	69	0

FREEDOM	37	154	7	16	9	361	0
GRIPHON	103	308	7	22	22	598	0
PATENT	19	56	0	2	6	144	10
SERAPHIN	60	217	10	21	13	399	0

[15]:
```
#match train and test columns
match_columns = list(X_train.columns)
# match_columns.remove('SV_EST')
# match_columns.remove('ALBUMIN_UNIT')
#match_columns.remove('REVEAL_BNP')
# match_columns.remove('diuretic_bi.1')
# print(match_columns)
X_test_match = X_test[match_columns]

missing_test_counts = pd.DataFrame({'test': X_test_match.isna().sum()/
 ↪len(X_test_match['SURVIVAL_1YR']),
                                    'train': X_train[match_columns].isna().sum()/
 ↪len(X_train['SURVIVAL_1YR'])})
pd.set_option('display.max_rows', 500)
missing_test_counts
```

[15]:

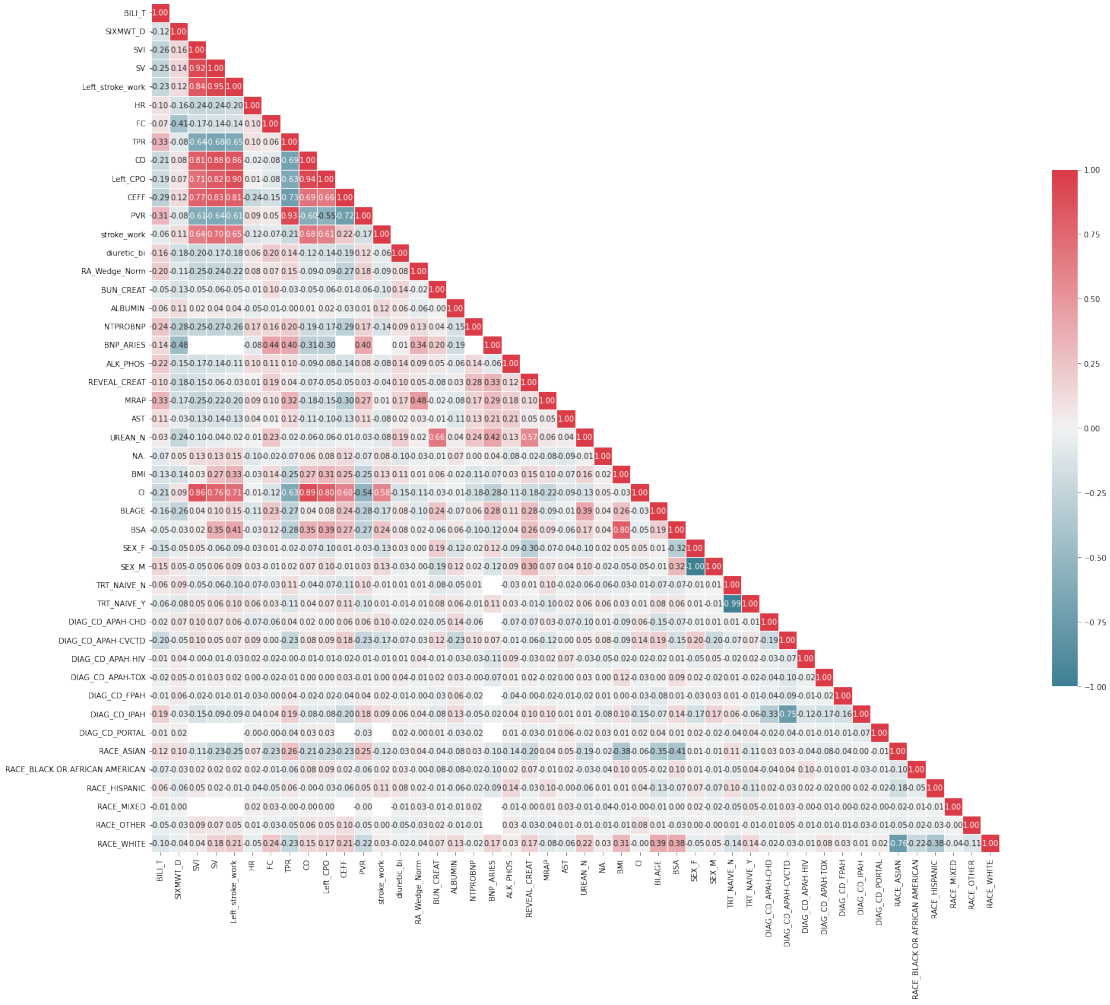
	test	train
USUBJID	0.000000	0.000000
BILI_T	0.013636	0.032439
REVEAL_BNP	0.028788	0.125350
SURVIVAL_1YR	0.000000	0.000000
SURVIVAL_1YRDY	0.000000	0.000000
DEATHFL_1_0	0.000000	0.000000
DEATHDY_FU	0.000000	0.000000
DEATHFL_FU	0.000000	0.000000
SIXMWT_D	0.000000	0.000000
CI	0.266667	0.161794
HR	0.000000	0.001602
MPAP	0.248485	0.146175
CO	0.266667	0.161794
PVR	0.266667	0.167401
PLT	0.072727	0.088506
SEX	0.000000	0.000000
FC	0.000000	0.000000
LEU	0.009091	0.058470
AST	0.007576	0.031237
ALK_PHOS	0.004545	0.030036
diuretic_bi	0.000000	0.000400
BLAGE	0.000000	0.000000
ACTARM	0.000000	0.000000

MRAP	0.304545	0.196636
PLACEBO	0.000000	0.000000

trial	0.000000	0.000000
diuretic_bi.1	0.000000	0.000400
HCT	0.007576	0.051662
HGB	0.006061	0.051662
RACE	0.000000	0.000400
DIAG_CD	0.000000	0.000000
REVEAL_CREAT	0.006061	0.029235
HEIGHT_CM	0.000000	0.000400
WEIGHT_KG	0.001515	0.001201
UREAN_N	0.031818	0.328394
NA.	0.004545	0.029235
SYSBP	0.000000	0.001201
DIABP	0.000000	0.001201
RAYNAUD	0.012121	0.004405
DIAG_DY	0.042424	0.056868
RBC	0.006061	0.051662
PCWP	0.278788	0.174610
RHC_HR	0.650000	0.496996
SIXMWT_BORG	0.001515	0.000400
TRT_NAIVE	0.012121	0.004405
era_bi	0.000000	0.000400
epoprostenol_bi	0.000000	0.000400
pde5_bi	0.000000	0.000400
pca_bi	0.000000	0.000400
riociguat_bi	0.000000	0.000400
NTPROBNP	0.080303	0.354425
BNP_ARIES	0.948485	0.971966
BNP_SERAPHIN	1.000000	0.798959
BSA	0.001515	0.001602
SV	0.653030	0.500601
SVI	0.653030	0.500601
stroke_work	0.653030	0.501001
stroke_work_index	0.653030	0.501001
Left_stroke_work	0.653030	0.501001
PVRI	0.268182	0.169003
RA_Wedge	0.331818	0.221466
TPR	0.266667	0.162195
CPO	0.266667	0.162195
Left_CPO	0.266667	0.162996
CPO_Index	0.268182	0.163797
CEFF	0.653030	0.501001
EGFR	0.006061	0.029636
ALBUMIN	0.004545	0.029636
ALBUMIN_UNIT	0.004545	0.029636
BMI	0.001515	0.001602

BUN_CREAT	0.033333	0.328394
RA_Wedge_Norm	0.331818	0.221466

[17]:
```
vars_only = X_train_continuous.drop(columns =
"SURVIVAL_1YR") from sklearn.preprocessing import StandardScaler
scaled_features = StandardScaler().fit_transform(vars_only) #standardized␣
↪dataset
scaled_features_df = pd.DataFrame(scaled_features) #changing dataset from array␣
↪to dataframe

import seaborn as sns #this is the library needed for the heat map
corr = scaled_features_df.corr() #computing the pearson correlation matrix for␣
↪your data
corr.columns = vars_only.columns
corr = corr.rename(index = lambda x: vars_only.columns[x])
# #if you need other correlation metrics, I'd suggest looking at the pandas␣
↪'DataFrame.corr' documentation


# # Generate a mask for the upper triangle
# #mask = np.triu(np.ones_like(corr, dtype=np.bool))
mask= np.triu(np.ones([len(corr),len(corr)])) #makes for a simplified figure
np.fill_diagonal(mask,0)

# # Set up the matplotlib figure
f, ax = plt.subplots(figsize=(25, 25))

# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True)

# Draw the heatmap with the mask and correct aspect ratio
sns_plot = sns.heatmap(corr, mask=mask, cmap=cmap, vmax=1, center=0,
            square=True, linewidths=.5, cbar_kws={"shrink": .5},annot=True,fmt="
↪2f")
plt.show()
```



[18]:
```
def treeCV(feature_column: np.ndarray, y: np.ndarray, max_leaf: int) ->␣
↪DecisionTreeClassifier:

    """Returns the best decisiontree"""

    rf_para = {'criterion':['entropy'],'max_leaf_nodes':[i for i in␣
↪range(2,max_leaf,1)],'min_samples_leaf':[100,200,250,300,1000]}
    clf = GridSearchCV(DecisionTreeClassifier(), rf_para, scoring =␣
↪'neg_brier_score')
    clf.fit(feature_column, y)

#     plot_tree(clf.best_estimator_)
    return clf
```

[109]:
```
from statsmodels.sandbox.stats.multicomp import multipletests
from scipy.stats import barnard_exact
```

[116]:
```
array([0, 0, 0, ..., 0, 1, 0])
```

[114]:
```
def significance_cutpoints(X,y) -> DecisionTreeClassifier:
    max_leaf = 5

    while max_leaf > 2:
#         if max_leaf < 2:
#             return cv_trees
        print(max_leaf)
        cv_trees = treeCV(X,y,max_leaf)
        variable_tree = cv_trees.best_estimator_
        variable_threshold = np.sort(variable_tree.tree_.threshold)
        variable_threshold = variable_threshold[~(variable_threshold == -2)]
        if len(variable_threshold) == 0:
            print("No cutpoints found.")
            return None
        X_cut = copy.deepcopy(X).flatten()
        max_val_train = np.max(X)
        label_names = np.linspace(1,len(variable_threshold)+1,num =
 ↪len(variable_threshold)+1, dtype = int)

        bins = [0,*variable_threshold,max_val_train]
#         print(X_cut)
        X_cut = pd.cut(X_cut,bins, right = True, labels = label_names)
#         all_combinations = list(combinations(label_names, 2))
#         print(all_combinations)
        all_combinations = []
        for i in label_names: #test p-value only for adjacent pairs
            if i + 1 in label_names:
                all_combinations.append((i,i+1))

        dummies = pd.get_dummies(X_cut, drop_first = True)
        all_p = []
#         print(dummies)
        for pair in all_combinations:
#             print(series)
            print(pair)
            c_xy = pd.crosstab(X_cut[(X_cut==pair[0]) |
 ↪(X_cut==pair[1])],y[(X_cut==pair[0]) | (X_cut==pair[1])])
#             print(c_xy)
            barnard_result = stats.barnard_exact(c_xy)
#             chi, p, _, _ = chi2_contingency(c_xy, correction = True)
#             print(chi,p)
            all_p.append(barnard_result.pvalue)
```

```
        _,corrected_p_vals,_,_ = multipletests(all_p, method="fdr_bh")
        print(corrected_p_vals)
#         print(any(corrected_p_vals > 0.05))
        if any(corrected_p_vals > 0.05):
            print("Chi-squared failed, reducing leaf node max")
            max_leaf = max_leaf - 1
            print(max_leaf)
        elif max_leaf == 2:
            return None
        else:
            print(cv_trees.best_estimator_)
            return cv_trees.best_estimator_


    return None
```

[120]:

```
best_trees = []
downselected_vars = ["stroke_work", "TPR", "CO", "Left_CPO","PVR","PVRI","CI",
                     "CEFF", "Left_stroke_work",␣
 ↪"stroke_work_index","SV","BNP_ARIES","BMI"]
# downselected_vars = ['BNP_ARIES']
#all_train_continuous=all_train_continuous.merge(all_train[['USUBJID',␣
 ↪'SURVIVAL_1YR']], on = 'USUBJID', how = 'left')
dummy_vars = ["SEX_F",
       "SEX_M", "TRT_NAIVE_N", "TRT_NAIVE_Y", "DIAG_CD_APAH-CHD",
       "DIAG_CD_APAH-CVCTD", "DIAG_CD_APAH-HIV", "DIAG_CD_APAH-TOX",
       "DIAG_CD_FPAH", "DIAG_CD_IPAH",␣
 ↪"DIAG_CD_PORTAL","diuretic_bi","RACE_ASIAN","RACE_HISPANIC",
             "RACE_MIXED", "RACE_OTHER", "RACE_WHITE"]

relevant_vars = []
for i in
    X_train_continuous.columns:
    print(i)
    if i in downselected_vars:
        print("removes")
        continue
    elif i in dummy_vars:
        continue
    elif i == "SURVIVAL_1YR":
        continue
    else:
#     print(i)

        match_cols = X_train_continuous[[i,"SURVIVAL_1YR"]]
        match_cols_valid = match_cols[~match_cols.isin([np.nan, np.inf, -np.
```

↪inf]
).any
(1)]

```
        predictor= np.array(match_cols_valid[i], ndmin = 2).
 ↪reshape(len(match_cols_valid),1)
#     print(predictor)
        y = np.array(match_cols_valid["SURVIVAL_1YR"],ndmin = 1)
        best_tree = significance_cutpoints(predictor,y)
        if i == "NA.":
            print(best_tree)
            plot_tree(best_tree)
            print(best_tree != None)
#             break
        if (best_tree != None):
            relevant_vars.append(i)
            best_trees.append(best_tree)
```

```
BILI_T
5
(1, 2)
(2, 3)
(3, 4)
[0.04378097 0.04378097 0.04378097]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=4,
                       min_samples_leaf=100)
SIXMWT_D
5
(1, 2)
(2, 3)
(3, 4)
[0.00109132 0.10012836 0.00109132]
Chi-squared failed, reducing leaf node  max
4
4
(1, 2)
(2, 3)
[0.0001678  0.00079087]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=3,
                       min_samples_leaf=250)
SVI
5
(1, 2)
[0.00013725]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=2,
                       min_samples_leaf=100)
SV
removes
Left_stroke_work
removes
```

HR
5

```
(1, 2)
(2, 3)
(3, 4)
[0.14052476 0.13040457 0.01872935]
Chi-squared failed, reducing leaf node  max
4
4
(1, 2)
(2, 3)
[0.02552041  0.01248623]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=3,
                       min_samples_leaf=100)
FC
5
(1, 2)
[0.00040427]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=2,
                       min_samples_leaf=100)
TPR
removes
CO
removes
Left_CPO
removes
CEFF
removes
PVR
removes
stroke_work
removes
diuretic_bi
RA_Wedge_Norm
5
(1, 2)
[0.00161055]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=2,
                       min_samples_leaf=100)
BUN_CREAT
5
No cutpoints found.
ALBUMIN
5
(1, 2)
(2, 3)
(3, 4)
[0.05618457 0.01247092 0.06025532]
Chi-squared failed, reducing leaf node  max
4
```

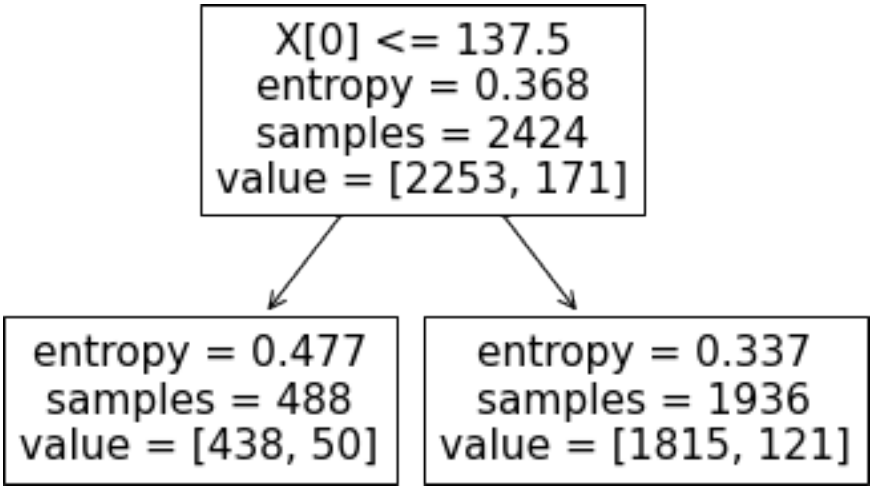
```
4
(1, 2)
(2, 3)
[0.03745638  0.00845337]
DecisionTreeClassifier(criterion='entropy', max_leaf_nodes=3,
                       min_samples_leaf=100)
SURVIVAL_1YR
NTPROBNP
5
(1, 2)
(2, 3)
[2.79975282e-02  3.31483690e-06]
DecisionTreeClassifier(criterion='entropy', max_leaf_nodes=3,
                       min_samples_leaf=300)
BNP_ARIES
removes
ALK_PHOS
5
(1, 2)
[0.00037799]
DecisionTreeClassifier(criterion='entropy', max_leaf_nodes=2,
                       min_samples_leaf=250)
REVEAL_CREAT
5
(1, 2)
(2, 3)
[0.02382547  0.00627313]
DecisionTreeClassifier(criterion='entropy', max_leaf_nodes=3,
                       min_samples_leaf=200)
MRAP
5
No cutpoints found.
AST
5
(1, 2)
(2, 3)
[0.05241477  0.02869787]
Chi-squared failed, reducing leaf node  max
4
4
(1, 2)
(2, 3)
[0.05241477  0.02869787]
Chi-squared failed, reducing leaf node  max
3
3
(1, 2)
```

[0.00177808]

```
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=2,
                       min_samples_leaf=100)
UREAN_N
5
(1, 2)
[0.00171293]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=2,
                       min_samples_leaf=300)
NA.
5
(1, 2)
(2, 3)
[0.05445118  0.07607703]
Chi-squared failed, reducing leaf node  max
4
4
(1, 2)
(2, 3)
[0.05445118  0.07607703]
Chi-squared failed, reducing leaf node  max
3
3
(1, 2)
[0.02853548]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=2,
                       min_samples_leaf=200)
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=2,
                       min_samples_leaf=200)
True
BMI
removes
CI
removes
BLAGE
5
(1, 2)
(2, 3)
[0.02673295  0.02673295]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=3,
                       min_samples_leaf=300)
BSA
5
(1, 2)
(2, 3)
(3, 4)
[0.03195425 0.03195425 0.01304759]
DecisionTreeClassifier(criterion="entropy", max_leaf_nodes=4,
```

```
min_samples_leaf=300)
```

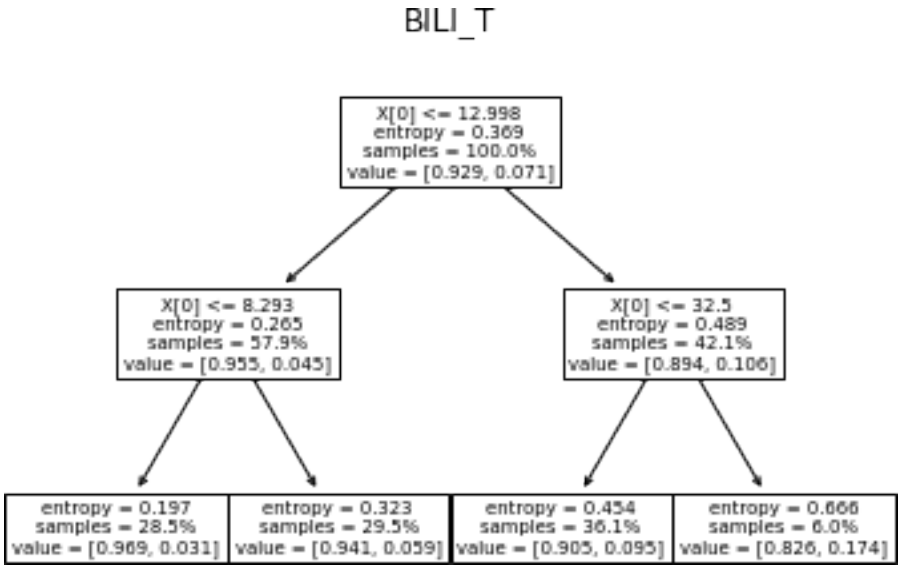
SEX_F
SEX_M
TRT_NAIVE_N
TRT_NAIVE_Y
DIAG_CD_APAH-CHD
DIAG_CD_APAH-CVCTD
DIAG_CD_APAH-HIV
DIAG_CD_APAH-TOX
DIAG_CD_FPAH
DIAG_CD_IPAH
DIAG_CD_PORTAL
RACE_ASIAN
RACE_BLACK OR AFRICAN AMERICAN
5
No cutpoints found.
RACE_HISPANIC
RACE_MIXED
RACE_OTHER
RACE_WHITE

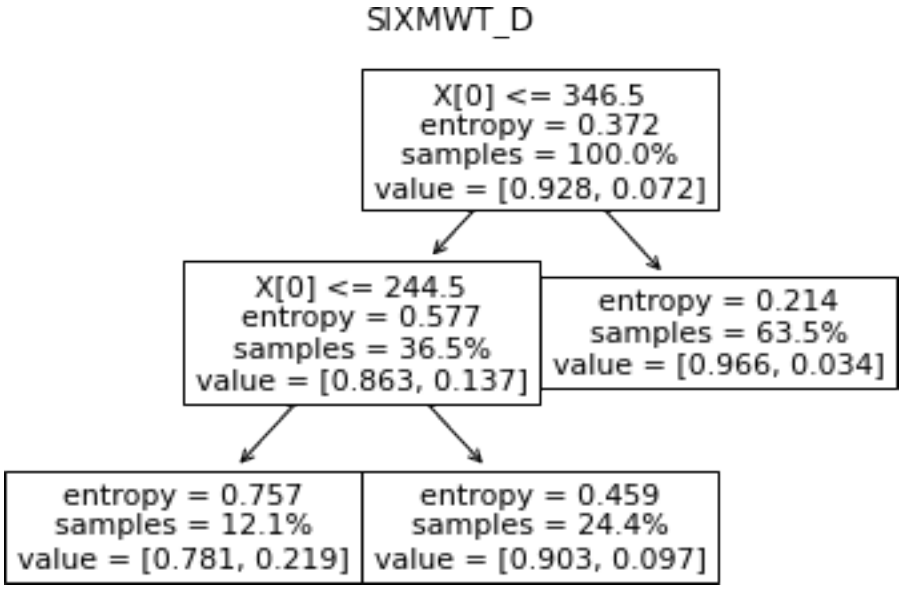


[121]:
```
pickle.dump(best_trees,open("dtrees_w_prune.p","wb"))
```

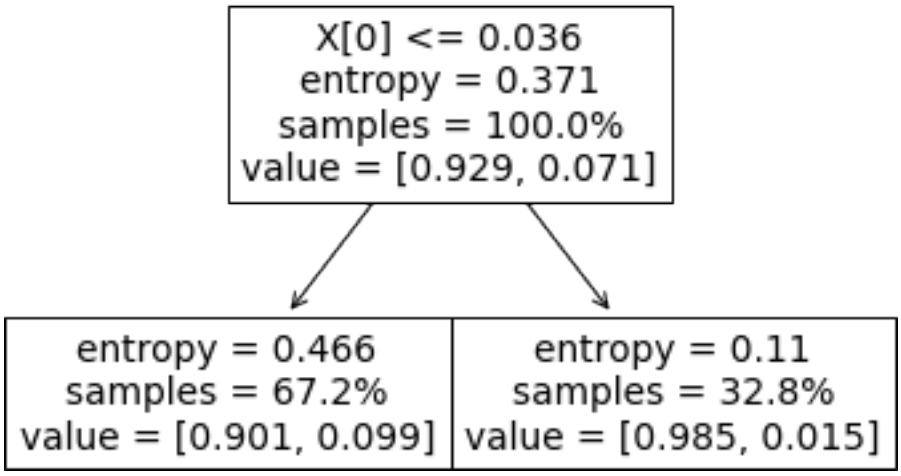
[122]:
```
for idx,i in  enumerate(relevant_vars):
#     if best_trees[idx] == None:
#         plt.title(i)
#     else:
    plot_tree(best_trees[idx], proportion = True)
```

```
    plt.title(i)
    plt.show()
```

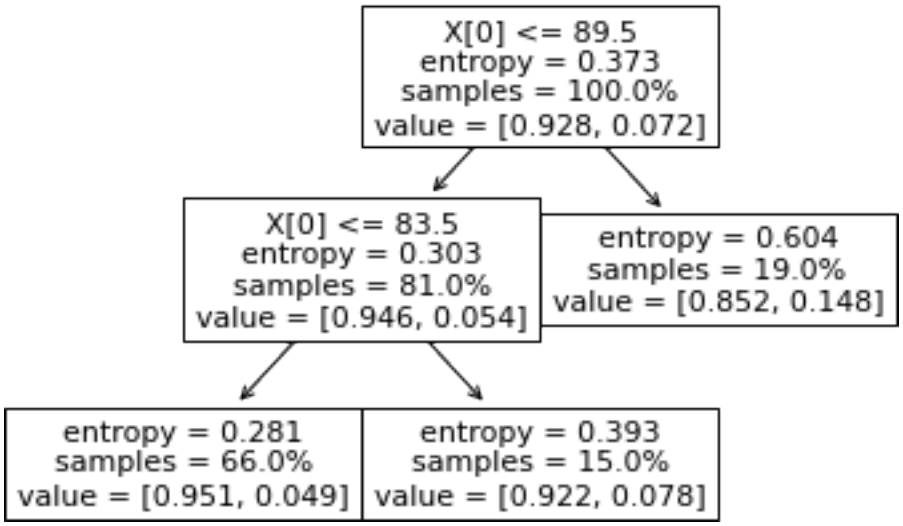




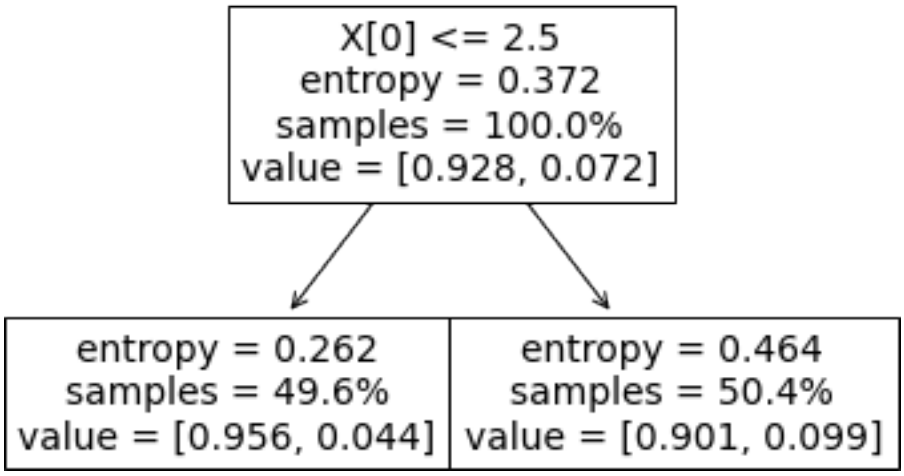
SVI


X[0] <= 0.036
entropy = 0.371
samples = 100.0%
value = [0.929, 0.071]
entropy = 0.466
samples = 67.2%
value = [0.901, 0.099]
entropy = 0.11
samples = 32.8%
value = [0.985, 0.015]

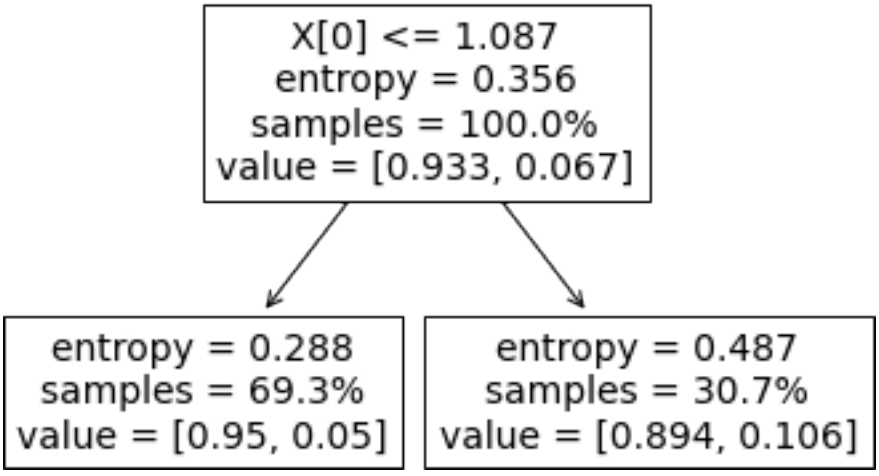
HR


X[0] <= 89.5
entropy = 0.373
samples = 100.0%
value = [0.928, 0.072]
X[0] <= 83.5
entropy = 0.303
samples = 81.0%
value = [0.946, 0.054]
entropy = 0.604
samples = 19.0%
value = [0.852, 0.148]
entropy = 0.281
samples = 66.0%
value = [0.951, 0.049]
entropy = 0.393
samples = 15.0%
value = [0.922, 0.078]

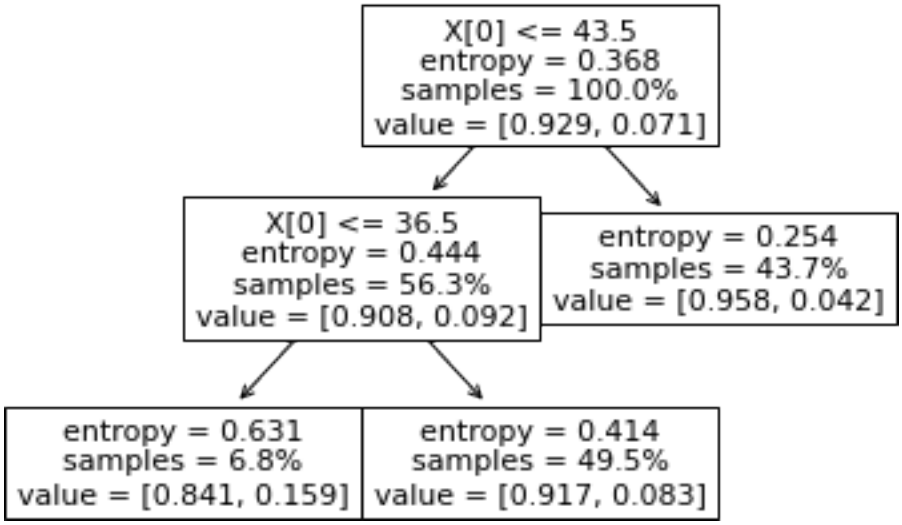
FC



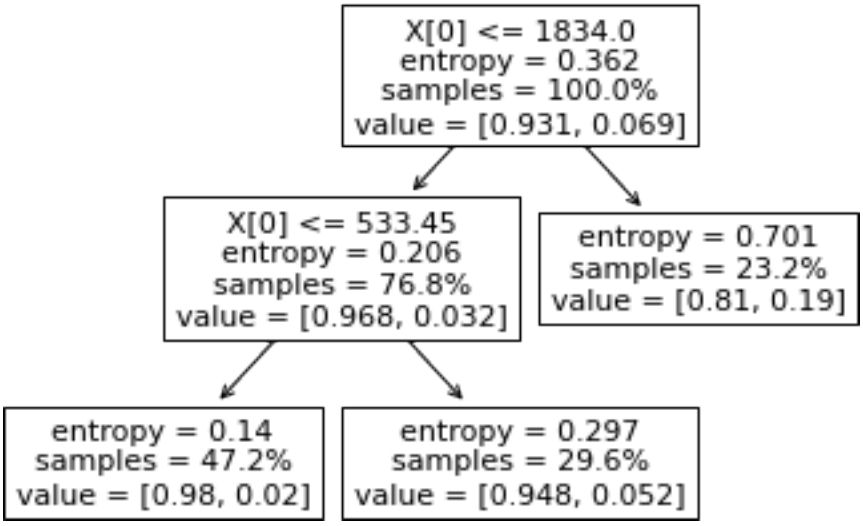
RA_Wedge_Norm



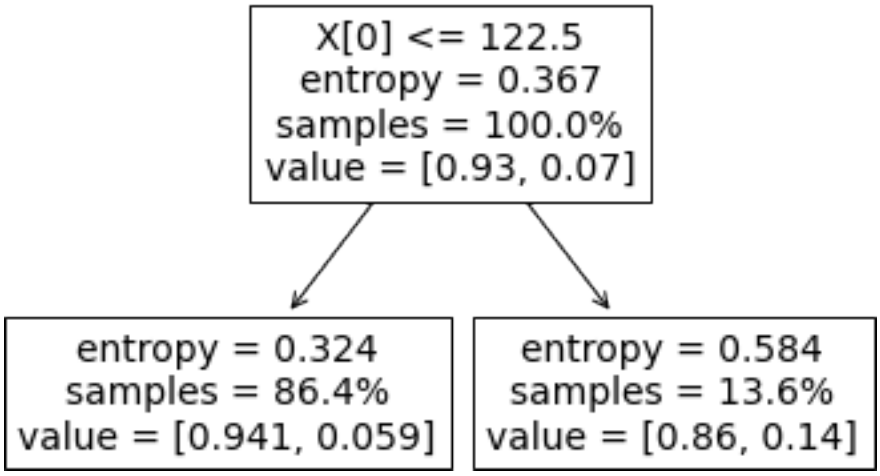
ALBUMIN



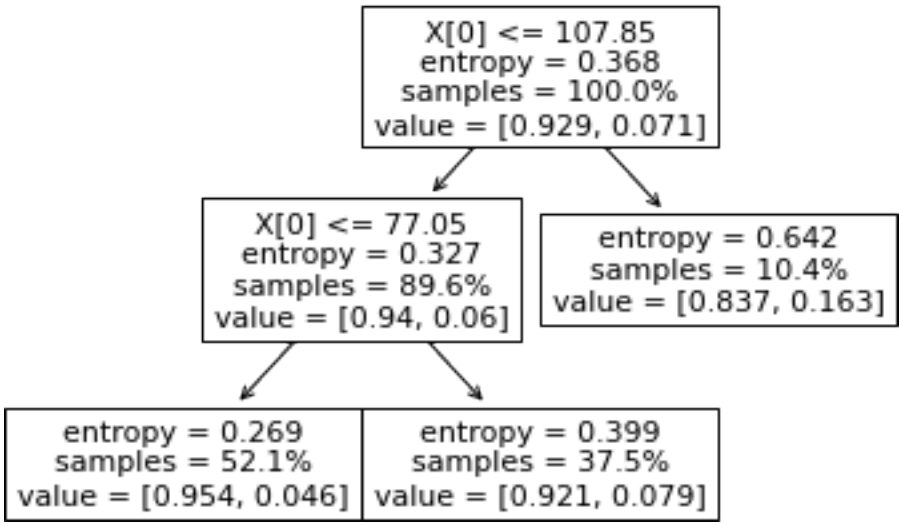
NTPROBNP



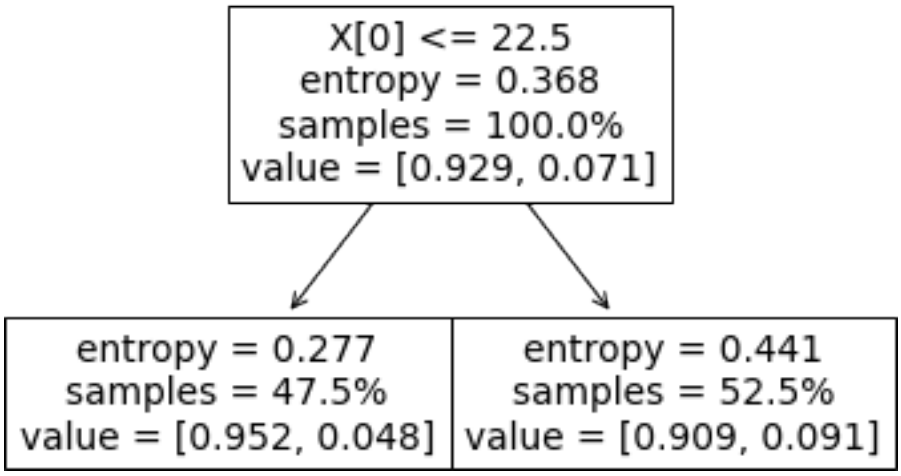
ALK_PHOS



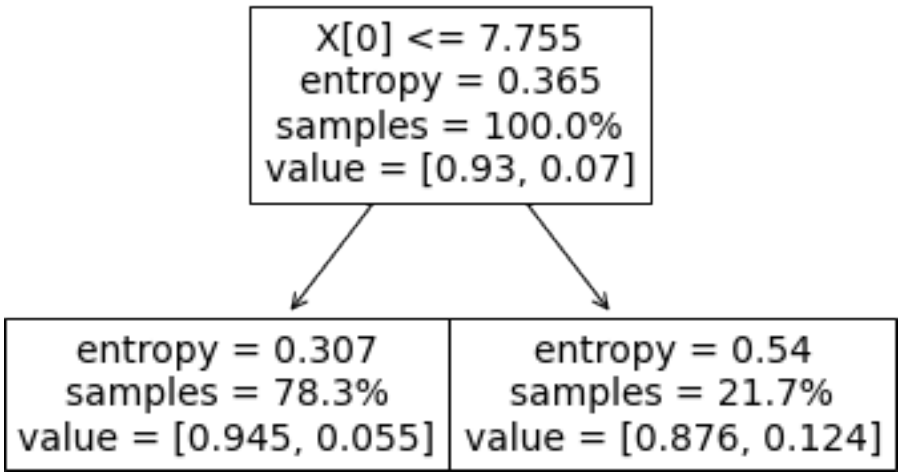
REVEAL_CREAT



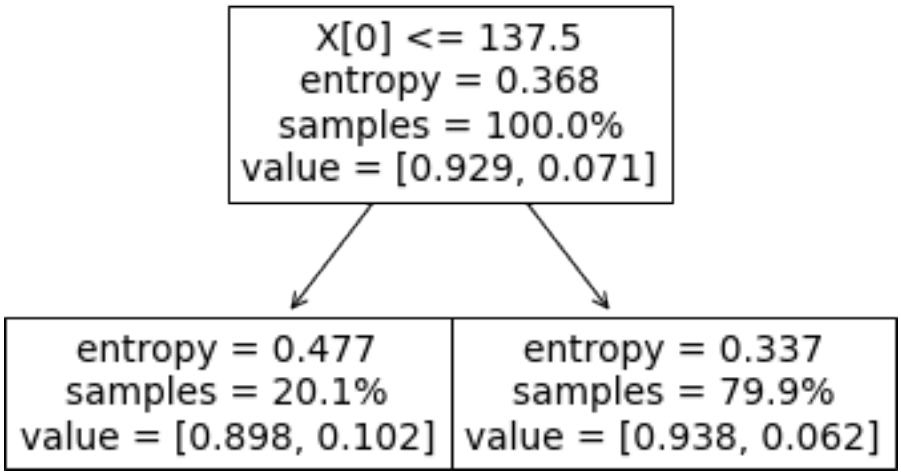
AST


X[0] <= 22.5
entropy = 0.368
samples = 100.0%
value = [0.929, 0.071]
entropy = 0.277
samples = 47.5%
value = [0.952, 0.048]
entropy = 0.441
samples = 52.5%
value = [0.909, 0.091]

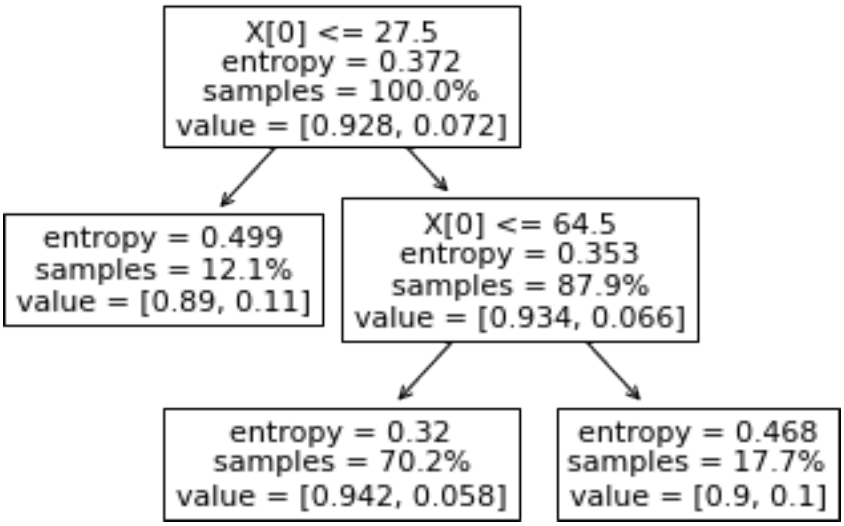
UREAN_N

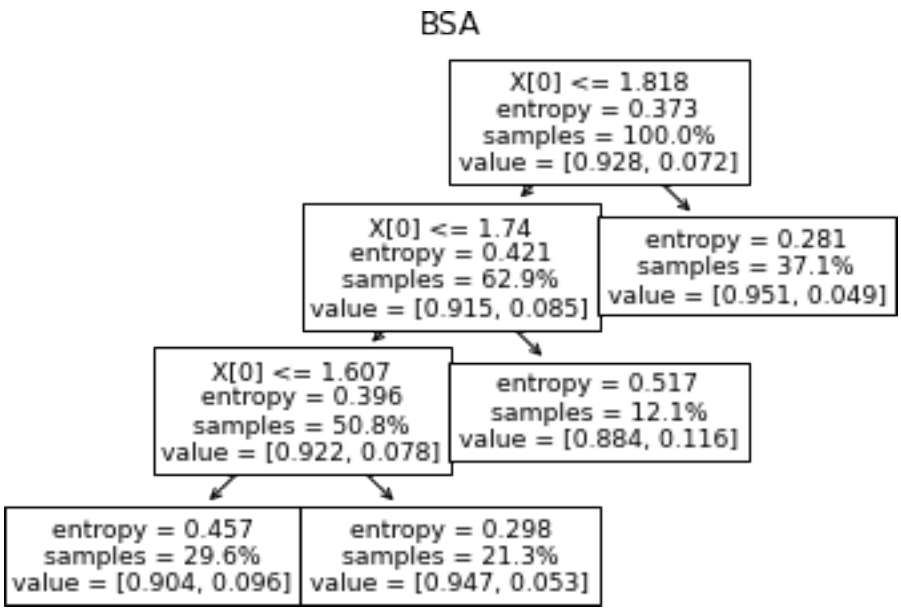

X[0] <= 7.755
entropy = 0.365
samples = 100.0%
value = [0.93, 0.07]
entropy = 0.307
samples = 78.3%
value = [0.945, 0.055]
entropy = 0.54
samples = 21.7%
value = [0.876, 0.124]

NA.


X[0] <= 137.5
entropy = 0.368
samples = 100.0%
value = [0.929, 0.071]
entropy = 0.477
samples = 20.1%
value = [0.898, 0.102]
entropy = 0.337
samples = 79.9%
value = [0.938, 0.062]

BLAGE


X[0] <= 27.5
entropy = 0.372
samples = 100.0%
value = [0.928, 0.072]
entropy = 0.499
samples = 12.1%
value = [0.89, 0.11]
X[0] <= 64.5
entropy = 0.353
samples = 87.9%
value = [0.934, 0.066]
entropy = 0.32
samples = 70.2%
value = [0.942, 0.058]
entropy = 0.468
samples = 17.7%
value = [0.9, 0.1]



[123]:
```
X_train_discretize = copy.deepcopy(X_train_continuous)
X_test_discretize =  copy.deepcopy(X_test)

X_test_discretize = pd.get_dummies(X_test_discretize, columns = discrete_vars)
# print(all_test_discretize)
dummy_vars = ["SEX_M", "DIAG_CD_APAH-CVCTD", "DIAG_CD_APAH-HIV",␣
 ↪"DIAG_CD_APAH-TOX", "DIAG_CD_FPAH",
             "DIAG_CD_IPAH", "DIAG_CD_PORTAL","DIAG_CD_UNKNOWN","TRT_NAIVE_Y",␣
 ↪"diuretic_bi","RACE_ASIAN"]

for idx,i in enumerate(relevant_vars + ["BNP_ARIES"]):
    print(i)
    if i in dummy_vars:
        continue
    max_val_train = np.max(X_train_continuous[i])
    max_val_test = np.max(X_test[i])
    max_val = np.max((max_val_train,max_val_test))
    if i == "BNP_ARIES":
        print(max_val)
        label_names = [1,2,3]
        bins = [0,49,180, max_val]
    elif i == "FC":
        continue
    else:
#         print(i)
```

```
        variable_tree = best_trees[idx]
        variable_threshold = np.sort(variable_tree.tree_.threshold)
        variable_threshold = variable_threshold[~(variable_threshold == -2)]
        label_names = np.linspace(1,len(variable_threshold)+1,num =␣
↪len(variable_threshold)+1, dtype = int)
        bins = [0,*variable_threshold,max_val]
    X_train_discretize[i] = pd.cut(X_train_discretize[i],bins, right = True,␣
↪labels = label_names)
    X_test_discretize[i] = pd.cut(X_test_discretize[i],bins, right = True,␣
↪labels = label_names)
X_train_discretize
```

```
BILI_T
SIXMWT_D
SVI
HR
FC
RA_Wedge_Norm
ALBUMIN
NTPROBNP
ALK_PHOS
REVEAL_CREAT
AST
UREAN_N
NA.
BLAGE
BSA
BNP_ARIES
2343.0
```

[123]:

	BILI_T	SIXMWT_D	SVI	SV	Left_stroke_work	HR	FC	TPR	CO	\
0	3	3	NaN	NaN	NaN	1	3	NaN	NaN	
1	3	3	NaN	NaN	NaN	1	2	NaN	NaN	
2	1	3	2	0.056579	3.998246	2	3	8.604651	4.30	
3	2	3	NaN	NaN	NaN	1	2	5.625000	8.00	
4	3	1	NaN	NaN	NaN	1	2	NaN	NaN	
...	...	...	...	...	...	..	..	...	...	
2492	NaN	3	NaN	NaN	NaN	2	3	7.121136	4.40	
2493	NaN	3	NaN	NaN	NaN	1	2	7.111111	4.50	
2494	NaN	3	NaN	NaN	NaN	1	3	13.495854	4.10	
2495	NaN	3	NaN	NaN	NaN	3	3	15.092736	4.13	
2496	NaN	3	NaN	NaN	NaN	1	3	13.671670	4.73	

	Left_CPO	...	DIAG_CD_APAH-TOX	DIAG_CD_FPAH	DIAG_CD_IPAH	\
0	NaN	...	0	0	1	
1	NaN	...	0	0	0	
2	0.673762	...	0	0	0	

```
3       1.300813  ...                   0                0                1
4            NaN  ...                   0                0                1
...          ...  ...                 ...              ...              ...
2492    0.773984  ...                   0                0                0
2493    0.831486  ...                   0                0                1
2494    0.787879  ...                   0                0                1
2495    0.805854  ...                   0                0                1
2496    0.873984  ...                   0                0                1

      DIAG_CD_PORTAL  RACE_ASIAN  RACE_BLACK OR AFRICAN AMERICAN  RACE_HISPANIC  \
0                  0           0                               0              0
1                  0           1                               0              0
2                  0           1                               0              0
3                  0           0                               0              1
4                  0           1                               0              0
...              ...         ...                             ...            ...
2492               0           0                               0              0
2493               0           0                               0              0
2494               0           0                               0              0
2495               0           1                               0              0
2496               0           0                               0              0

      RACE_MIXED  RACE_OTHER  RACE_WHITE
0              0           0           1
1              0           0           0
2              0           0           0
3              0           0           0
4              0           0           0
...          ...         ...         ...
2492           0           0           1
2493           0           0           1
2494           0           0           1
2495           0           0           0
2496           0           0           1

[2497 rows x 47 columns]
```

[124]:
```
##Combine BNP and NTproBNP
X_test_discretize["BNP_Combined"] = np.
↪nansum([X_test_discretize["BNP_ARIES"],X_test_discretize["NTPROBNP"]], axis =␣
↪0)
X_test_discretize["BNP_Combined"] = X_test_discretize["BNP_Combined"].replace(0,␣
↪np.nan, inplace=False)

X_train_discretize["BNP_Combined"] = np.
↪nansum([X_train_discretize["BNP_ARIES"],X_train_discretize["NTPROBNP"]], axis␣
↪= 0)
```

```
X_train_discretize["BNP_Combined"] = X_train_discretize["BNP_Combined"].
↪replace(0, np.nan, inplace=False)
```

[167]:
```
stats_outcomes = []

untested_vars = ["SEX_F",
       "SEX_M", "TRT_NAIVE_N", "TRT_NAIVE_Y", "DIAG_CD_APAH-CHD",
       "DIAG_CD_APAH-CVCTD", "DIAG_CD_APAH-HIV", "DIAG_CD_APAH-TOX",
       "DIAG_CD_FPAH", "DIAG_CD_IPAH",␣
↪"DIAG_CD_PORTAL","RACE_ASIAN","diuretic_bi"] + ["BNP_Combined","FC"]
relevant_vars2 = []

for var in untested_vars:
    if var == "DIAG_CD_UNKNOWN":
        continue
    label_names = np.unique(X_train_discretize[var])[~np.isnan(np.
↪unique(X_train_discretize[var]))]
#     print(label_names)
#     if var == 'BNP_Combined':
#         continue
    all_combinations = list(combinations(label_names, 2))
    all_p = []
    X_var = X_train_discretize[var]
    y = X_train_discretize["SURVIVAL_1YR"]
#     print(all_combinations)
    for pair in all_combinations:
        print(pair)
        c_xy = pd.crosstab(X_var[(X_var==pair[0]) | (X_var==pair[1])],
                           y[(X_var==pair[0]) | (X_var==pair[1])])
        barnard_result = stats.barnard_exact(c_xy)
#             chi, p, _, _ = chi2_contingency(c_xy, correction = True)
#             print(chi,p)
#         chi, p, _, _ = chi2_contingency(c_xy, correction = True)

        all_p.append(barnard_result.pvalue)
        _,corrected_p_vals,_,_ = multipletests(all_p, method="fdr_bh")
    print(var)
    print(corrected_p_vals)
    if all(corrected_p_vals <
        0.05):
        relevant_vars2.append(var)
```

```
(0, 1)
SEX_F
[0.00183837]
(0, 1)
```

SEX_M
[0.00183837]
(0, 1)

TRT_NAIVE_N
[0.28855421]
(0, 1)
TRT_NAIVE_Y
[0.58980431]
(0, 1)
DIAG_CD_APAH-CHD
[0.89163578]
(0, 1) DIAG_CD_APAH-
CVCTD [0.01982959]
(0, 1)
DIAG_CD_APAH-
HIV [1.]
(0, 1)
DIAG_CD_APAH-TOX
[0.98132765]
(0, 1)
DIAG_CD_FPAH
[0.0571313]
(0, 1)
DIAG_CD_IPAH
[0.10571307]
(0, 1)
DIAG_CD_PORTAL
[0.74023087]
(0, 1)
RACE_ASIAN
[0.4019599]
(0.0, 1.0)
diuretic_bi
[0.0288797]
(1.0, 2.0)
(1.0, 3.0)
(2.0, 3.0)
BNP_Combined
[1.33253675e-02 4.72625163e-12 1.89990876e-06]
(1, 2)
(1, 3)
(1, 4)
(2, 3)
(2, 4)
(3, 4)
FC
[3.65565962e-01 1.41417412e-01 1.46906170e-02 4.52966266e-03
 2.55401071e-04 6.45939971e-03]

[ ]:
```
for col in final_test:
    print(col)
    print(final_test[col].unique())
```

[133]:
```
from sklearn import metrics, preprocessing
from scipy import stats


x = final_train
n = len(x.columns)
matMI = np.zeros((n, n))

# print(x)
for ix in np.arange(n):

    for jx in np.arange(n):
#         print(x.columns[ix],x.columns[jx])
#         xy = x[[x.columns[ix],x.columns[jx]]]
#         print(xy)
#         xy = xy.dropna()

#         first = np.array(xy[[x.columns[ix]]]).flatten()
#         second = np.array(xy[[x.columns[jx]]]).flatten()
        ix_name = x.columns[ix]
        jx_name = x.columns[jx]
        test_corr = final_train[[ix_name,jx_name,"SURVIVAL_1YR"]]
        test_corr = test_corr.dropna()
        test_corr = test_corr.loc[:,~test_corr.columns.duplicated()]
#         print(test_corr)
        test_corr["ix_Predict"] = test_corr.groupby(ix_name)["SURVIVAL_1YR"].
↪transform("mean")
        test_corr["jx_Predict"] = test_corr.groupby(jx_name)["SURVIVAL_1YR"].
↪transform("mean")
#         print(test_corr)
        concordance = kendalltau(test_corr["ix_Predict"],test_corr["jx_Predict"])
#         print(concordance)
#         mis = metrics.normalized_mutual_info_score(first,second)
#         matMI[ix,jx] = mis
        matMI[ix,jx] = concordance[0]
#         break
```

[135]:
```
print(x.columns)
```

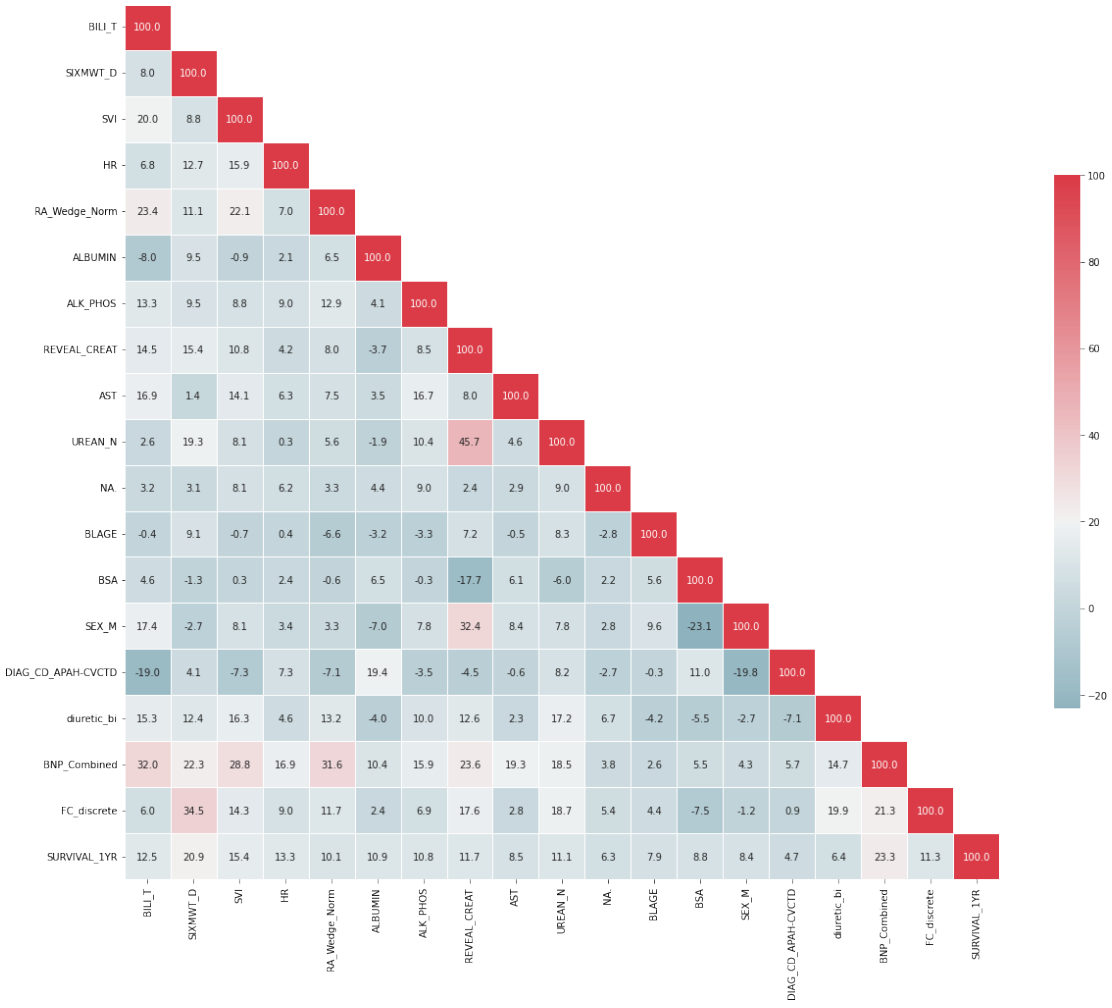
```
Index(["BILI_T", "SIXMWT_D", "SVI", "HR", "RA_Wedge_Norm", "ALBUMIN",
       "ALK_PHOS", "REVEAL_CREAT", "AST", "UREAN_N", "NA.", "BLAGE", "BSA",
```

```
       ined', 'SURVIVAL_1YR', 'FC_discrete'],
      dtype='object')
```

```
'SEX_M';'DIAG_CD_APAH-CVCTD';'diuretic_bi';'BNP_Comb
```

[136]:
```
import matplotlib.pyplot as plt

import seaborn as  sns

#mask= np.triu(np.ones([len(matMI),len(matMI)])) #makes for a simplified figure
#np.fill_diagonal(mask,0)
#np.fill_diagonal(matMI,1)

# i_lower = np.tril_indices(len(matMI), -1)
# i_upper = np.triu_indices(len(matMI), 1)
# matMI = np.rot90(np.fliplr(matMI))
# matMI_new[i_lower] = matMI_new[i_upper]

mask= np.triu(np.ones([len(x.columns),len(x.columns)])) #makes for a simplified␣
 ↪figure
np.fill_diagonal(mask,0)

# Set up the matplotlib figure
f, ax = plt.subplots(figsize=(20, 20))

# Generate a custom diverging colormap
cmap = sns.diverging_palette(220, 10, as_cmap=True)
matMIdf = pd.DataFrame(matMI, columns = x.columns, index = x.columns)
matMIdf = matMIdf.multiply(100)
# print(matMIdf.max())
# all_but_survival = matMIdf.columns

matMIdf = matMIdf.reindex(['BILI_T', 'SIXMWT_D', 'SVI', 'HR', 'RA_Wedge_Norm',␣
 ↪'ALBUMIN',
       'ALK_PHOS', 'REVEAL_CREAT', 'AST', 'UREAN_N', 'NA.', 'BLAGE', 'BSA',
       'SEX_M', 'DIAG_CD_APAH-CVCTD', 'diuretic_bi',␣
 ↪'BNP_Combined','FC_discrete','SURVIVAL_1YR'])

matMIdf = matMIdf[['BILI_T', 'SIXMWT_D', 'SVI', 'HR', 'RA_Wedge_Norm', 'ALBUMIN',
       'ALK_PHOS', 'REVEAL_CREAT', 'AST', 'UREAN_N', 'NA.', 'BLAGE', 'BSA',
       'SEX_M', 'DIAG_CD_APAH-CVCTD', 'diuretic_bi',␣
 ↪'BNP_Combined','FC_discrete','SURVIVAL_1YR']]

# Draw the heatmap with the mask and correct aspect ratio
sns.heatmap(matMIdf,cmap = cmap,mask = mask, vmax=matMIdf.max().max(), vmin =␣
↪matMIdf.min().min(), center=20,
            square=True, linewidths=.5, cbar_kws={"shrink": .5},annot=True,fmt=".
 ↪1f")
```

[136]: <AxesSubplot:>

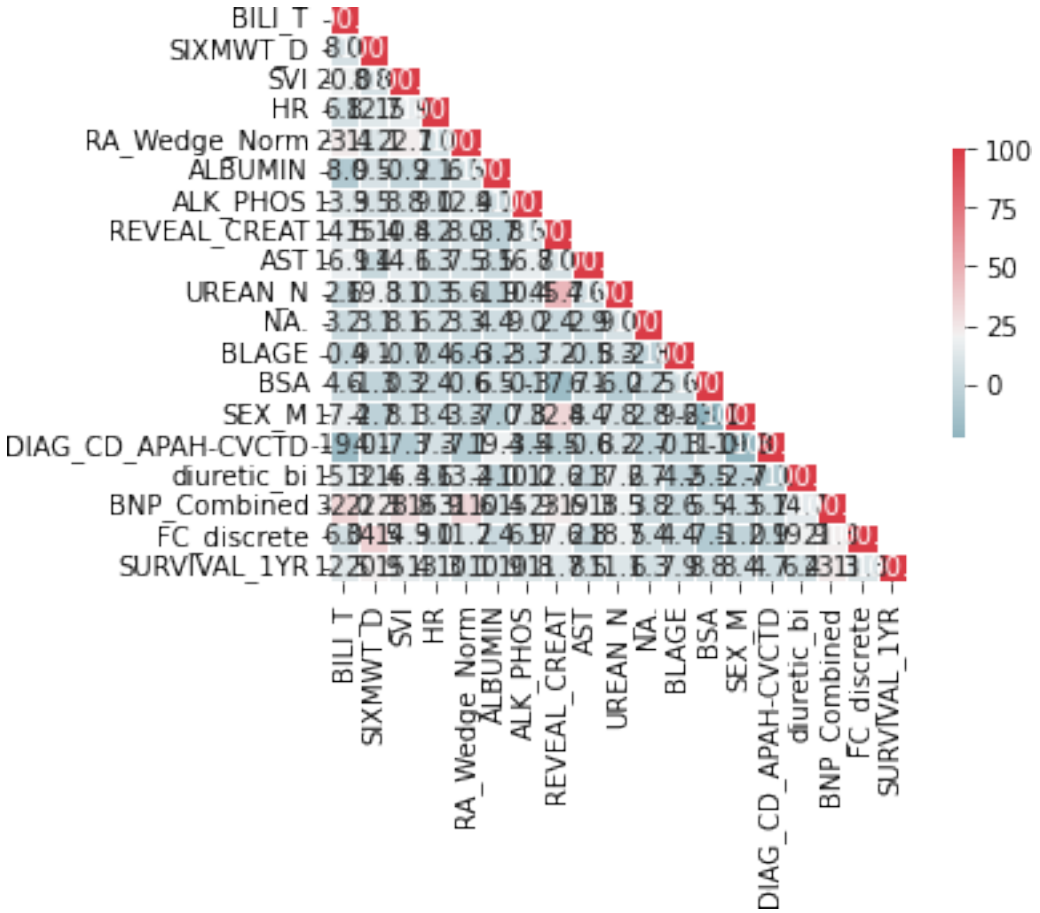


[ ]:
```
x = matMIdf.sort_values(by = ["SURVIVAL_1YR"])
```

[77]:
```
print(x)
```

['BILI_T', 'SIXMWT_D', 'SVI', 'HR', 'RA_Wedge_Norm', 'ALBUMIN', 'ALK_PHOS', 'REVEAL_CREAT', 'AST', 'UREAN_N', 'NA.', 'BLAGE', 'BSA', 'SEX_M', 'DIAG_CD_APAH-CVCTD', 'diuretic_bi', 'BNP_Combined', 'FC_discrete']

[137]:
```
heatmap = sns.heatmap(matMIdf,cmap = cmap,mask = mask, vmax=matMIdf.max().max(),␣
↪vmin = matMIdf.min().min(), center=20,
            square=True, linewidths=.5, cbar_kws={"shrink": .5},annot=True,fmt=".
↪1f")
fig = heatmap.get_figure()
fig.savefig("kendall_tau_matrix_prune.png")
```



[ ]:
```
# x = final_train


# ix_name = 'DIAG_CD_APAH-CVCTD'
# jx_name = 'SEX_M'
# test_corr = final_train[[ix_name,jx_name,'SURVIVAL_1YR']]
# test_corr = test_corr.dropna()
# test_corr = test_corr.loc[:,~test_corr.columns.duplicated()]
# #         print(test_corr)
# test_corr['Connective_Tissue_Disorder'] = test_corr.
 ↪groupby(ix_name)['SURVIVAL_1YR'].transform('mean')
# test_corr['Male_Sex'] = test_corr.groupby(jx_name)['SURVIVAL_1YR'].
 ↪transform('mean')
# # test_corr
# pd.crosstab(test_corr['Connective_Tissue_Disorder'],test_corr['Male_Sex']).
```

[ ]:
```
# x=matMIdf[matMIdf.columns[[0,1]]]
# x.loc['SURVIVAL_1YR']
# all_combinations = list(combinations(['BILI_T','SIXMWT_D','BNP_Combined'], 2))
# print(all_combinations)
# print(all_combinations[0])
# which_column = matMIdf[[all_combinations[0][0]]]
# print(which_column)
# print(all_combinations[0][1])
# which_row = which_column.loc[all_combinations[0][1],:].values[0]
# print(which_row)
```

[49]:
```
from itertools import chain, combinations
# def all_subsets(ss):
#     return chain(*map(lambda x: combinations(ss, x), range(0, len(ss)+1)))

# for subset in all_subsets(matMIdf.columns):
#     print(subset)
```

[138]:
```
matMIdf_feat_select = matMIdf.drop(columns = "SURVIVAL_1YR")
matMIdf_feat_select
```

[138]:

	BILI_T	SIXMWT_D	SVI	HR	\
BILI_T	100.000000	7.968168	20.003820	6.779069	
SIXMWT_D	7.968168	100.000000	8.812329	12.717755	
SVI	20.003820	8.812329	100.000000	15.937344	
HR	6.779069	12.717755	15.937344	100.000000	
RA_Wedge_Norm	23.382964	11.139013	22.119216	6.998191	
ALBUMIN	-8.029553	9.505406	-0.855073	2.117831	
ALK_PHOS	13.292643	9.462048	8.811028	9.032969	
REVEAL_CREAT	14.549133	15.408418	10.827471	4.160649	
AST	16.906554	1.437547	14.068678	6.271946	
UREAN_N	2.561807	19.311811	8.111584	0.260839	
NA.	3.176001	3.082257	8.062001	6.150500	
BLAGE	-0.353484	9.125273	-0.687853	0.375215	
BSA	4.579512	-1.297607	0.260830	2.353979	
SEX_M	17.386909	-2.739707	8.119954	3.393224	
DIAG_CD_APAH-CVCTD	-18.993636	4.132805	-7.275227	7.328279	
diuretic_bi	15.272347	12.423012	16.343821	4.630243	
BNP_Combined	31.994075	22.266733	28.842747	16.932158	
FC_discrete	6.033015	34.531383	14.347559	9.026346	
SURVIVAL_1YR	12.547893	20.853530	15.386871	13.314821	

	RA_Wedge_Norm	ALBUMIN	ALK_PHOS	REVEAL_CREAT	\
BILI_T	23.382964	-8.029553	13.292643	14.549133	
SIXMWT_D	11.139013	9.505406	9.462048	15.408418	

SVI	22.119216	-0.855073	8.811028	10.827471
HR	6.998191	2.117831	9.032969	4.160649

RA_Wedge_Norm	100.000000	6.502008	12.913613	7.973676
ALBUMIN	6.502008	100.000000	4.130201	-3.670102
ALK_PHOS	12.913613	4.130201	100.000000	8.527164
REVEAL_CREAT	7.973676	-3.670102	8.527164	100.000000
AST	7.509010	3.532014	16.705813	8.034594
UREAN_N	5.625924	-1.862814	10.412028	45.666147
NA.	3.345667	4.417395	8.973972	2.356711
BLAGE	-6.628649	-3.185189	-3.307052	7.173845
BSA	-0.562273	6.521738	-0.300068	-17.667333
SEX_M	3.262483	-7.007197	7.759431	32.376757
DIAG_CD_APAH-CVCTD	-7.142383	19.378947	-3.537096	-4.518575
diuretic_bi	13.213647	-4.029151	10.025348	12.571538
BNP_Combined	31.622157	10.395713	15.915001	23.562772
FC_discrete	11.694638	2.441245	6.919575	17.638118
SURVIVAL_1YR	10.060549	10.924882	10.805899	11.709720

	AST	UREAN_N	NA.	BLAGE \
BILI_T	16.906554	2.561807	3.176001	-0.353484
SIXMWT_D	1.437547	19.311811	3.082257	9.125273
SVI	14.068678	8.111584	8.062001	-0.687853
HR	6.271946	0.260839	6.150500	0.375215
RA_Wedge_Norm	7.509010	5.625924	3.345667	-6.628649
ALBUMIN	3.532014	-1.862814	4.417395	-3.185189
ALK_PHOS	16.705813	10.412028	8.973972	-3.307052
REVEAL_CREAT	8.034594	45.666147	2.356711	7.173845
AST	100.000000	4.577600	2.922265	-0.496184
UREAN_N	4.577600	100.000000	8.987598	8.271393
NA.	2.922265	8.987598	100.000000	-2.786529
BLAGE	-0.496184	8.271393	-2.786529	100.000000
BSA	6.092459	-6.027063	2.239173	5.612340
SEX_M	8.351029	7.756380	2.771861	9.636594
DIAG_CD_APAH-CVCTD	-0.575176	8.174267	-2.719585	-0.298636
diuretic_bi	2.286591	17.161302	6.704359	-4.189029
BNP_Combined	19.338226	18.460565	3.849490	2.560485
FC_discrete	2.781397	18.674704	5.410779	4.405832
SURVIVAL_1YR	8.492104	11.131505	6.257469	7.851861

	BSA	SEX_M	DIAG_CD_APAH-CVCTD	diuretic_bi \
BILI_T	4.579512	17.386909	-18.993636	15.272347
SIXMWT_D	-1.297607	-2.739707	4.132805	12.423012
SVI	0.260830	8.119954	-7.275227	16.343821
HR	2.353979	3.393224	7.328279	4.630243
RA_Wedge_Norm	-0.562273	3.262483	-7.142383	13.213647
ALBUMIN	6.521738	-7.007197	19.378947	-4.029151

ALK_PHOS	-0.300068	7.759431	-3.537096	10.025348
REVEAL_CREAT	-17.667333	32.376757	-4.518575	12.571538
AST	6.092459	8.351029	-0.575176	2.286591

UREAN_N	-6.027063	7.756380	8.174267	17.161302
NA.	2.239173	2.771861	-2.719585	6.704359
BLAGE	5.612340	9.636594	-0.298636	-4.189029
BSA	100.000000	-23.148732	10.991939	-5.491742
SEX_M	-23.148732	100.000000	-19.845163	-2.732984
DIAG_CD_APAH-CVCTD	10.991939	-19.845163	100.000000	-7.080097
diuretic_bi	-5.491742	-2.732984	-7.080097	100.000000
BNP_Combined	5.478652	4.277415	5.670104	14.717802
FC_discrete	-7.504986	-1.180403	0.895353	19.861399
SURVIVAL_1YR	8.821842	8.372487	4.664692	6.369760

	BNP_Combined	FC_discrete
BILI_T	31.994075	6.033015
SIXMWT_D	22.266733	34.531383
SVI	28.842747	14.347559
HR	16.932158	9.026346
RA_Wedge_Norm	31.622157	11.694638
ALBUMIN	10.395713	2.441245
ALK_PHOS	15.915001	6.919575
REVEAL_CREAT	23.562772	17.638118
AST	19.338226	2.781397
UREAN_N	18.460565	18.674704
NA.	3.849490	5.410779
BLAGE	2.560485	4.405832
BSA	5.478652	-7.504986
SEX_M	4.277415	-1.180403
DIAG_CD_APAH-CVCTD	5.670104	0.895353
diuretic_bi	14.717802	19.861399
BNP_Combined	100.000000	21.295909
FC_discrete	21.295909	100.000000
SURVIVAL_1YR	23.308811	11.332621

[51]:
```
from itertools import combinations
from scipy.optimize import differential_evolution, rosen, minimize
```

[177]:
```
def redundancy_cost(parameterTuple):
    feat_select = (parameterTuple > 0.9)
    nonzero_vars = np.argwhere(feat_select).flatten()     #find which variables
↪are contributing
#     print(matMldf_feat_select.columns[[nonzero_vars]])
    positive_contribution = matMldf_feat_select[matMldf_feat_select.
↪columns[[nonzero_vars]]]
    total_positive = np.sum(positive_contribution.loc["SURVIVAL_1YR"].values)
    all_combinations = list(combinations(matMldf_feat_select.
↪columns[[nonzero_vars]], 2))#get combinations
    total_negative = 0
    for pair in all_combinations:
```

```
            which_column = matMldf_feat_select[[pair[0]]]
            which_row = which_column.loc[pair[1],:].values[0]
            total_negative = total_negative + which_row
        noninteger_penalty = np.sum((parameterTuple - np.round(parameterTuple))**2)
        return -total_positive + 0.125*total_negative + 10*noninteger_penalty

def printCurrentIteration(xk, convergence):
        print("finished iteration")
        print(xk)
        print(convergence)

def genetic_search():
        parameterBounds = []
        for i in range(len(matMldf_feat_select)-1):
#             if i == 1:
#                 parameterBounds.append([1,1])# six minute walk distance
#             elif i == 2:
#                 parameterBounds.append([1,1])# svi
#             elif i == 16:
#                 parameterBounds.append([1,1])# all other variables
#             else:
            parameterBounds.append([0,1])# all other variables

        result = differential_evolution(redundancy_cost, parameterBounds, seed=3,
    ↪maxiter =100)

        return result.x


geneticParameters = genetic_search()
```

[178]:
```
# print(np.round(geneticParameters))
print(geneticParameters)
# print(x)
# print(geneticParameters>.1)
```

```
[0.9691162  0.97250292 0.93131277 0.90383983 0.84147106 0.96600705
 0.9380102  0.15643325 0.89854685 0.11849082 0.99695734 0.98533046
 0.91341792 0.93372354 0.97112605 0.00470437 0.95193661 0.7984779 ]
```

[179]:

[180]:

```
print(np.asarray(matMldf_feat_select.columns)[(geneticParameters < .9)])
```

```
['RA_Wedge_Norm' 'REVEAL_CREAT' 'AST' 'UREAN_N' 'diure
```

```
chosen_features = list(matMldf_feat_select.columns)
```

```
print(np.asarray(chosen_features)[(geneticParameters >  .9)])
# print(np.asarray(x)[~(np.round(geneticParameters) == 1)])
selected_features = np.asarray(chosen_features)[(geneticParameters >  .9)]
pickle.dump(selected_features, open("penalty_pt125_10_8_prune","wb"))
```

```
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'ALK_PHOS' 'NA.' 'BLAGE' 'BSA'
 'SEX_M' 'DIAG_CD_APAH-CVCTD' 'BNP_Combined']
```

[164]:
```
import pickle
pt05 = pickle.load(open("penalty_pt05_10_8_prune","rb"))
pt075 =
pickle.load(open("penalty_pt075_10_8_prune","rb")) pt1 =
pickle.load(open("penalty_pt1_10_8_prune","rb")) pt15 =
pickle.load(open("penalty_pt15_10_8_prune","rb")) pt175 =
pickle.load(open("penalty_pt175_10_8_prune","rb")) pt2 =
pickle.load(open("penalty_pt2_10_8_prune","rb"))
```

[168]:
```
print(pt05)
print(pt075)
print(pt1)
print(pt15)
print(pt175)
print(pt2)
```

```
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'ALK_PHOS' 'REVEAL_CREAT' 'AST'
 'UREAN_N' 'BLAGE' 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD' 'diuretic_bi'
 'BNP_Combined' 'FC_discrete']
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'RA_Wedge_Norm' 'ALBUMIN' 'ALK_PHOS'
 'UREAN_N' 'NA.' 'BLAGE' 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD' 'BNP_Combined'
 'FC_discrete']
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'ALK_PHOS' 'UREAN_N' 'NA.'
 'BLAGE' 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD' 'BNP_Combined']
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'ALK_PHOS' 'UREAN_N' 'NA.'
 'BLAGE' 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD']
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'ALK_PHOS' 'UREAN_N' 'BLAGE'
 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD']
['SIXMWT_D' 'HR' 'ALBUMIN' 'NA.' 'BLAGE' 'BSA' 'SEX_M'
 'DIAG_CD_APAH-CVCTD' 'diuretic_bi' 'BNP_Combined']
```

[83]:
```
import pickle
pt05 = pickle.load(open("penalty_pt05_10_8","rb"))
pt075 = pickle.load(open("penalty_pt075_10_8","rb"))
pt1 = pickle.load(open("penalty_pt1_10_7","rb"))
pt15 = pickle.load(open("penalty_pt15_10_7","rb"))
pt2 = pickle.load(open("penalty_pt2_10_7","rb"))
pt3 = pickle.load(open("penalty_pt3_10_7","rb"))
pt4 = pickle.load(open("penalty_pt4_10_7","rb"))
# print(pt10)
```

```
print(pt1)
print(pt15)
print(pt2)
print(pt3)
print(pt4)
```

```
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'ALK_PHOS'
 'UREAN_N' 'NA.' 'BLAGE' 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD'
 'BNP_Combined']
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'ALK_PHOS' 'REVEAL_CREAT'
 'NA.' 'BLAGE' 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD' 'BNP_Combined']
['BILI_T' 'SIXMWT_D' 'SVI' 'HR' 'ALBUMIN' 'NA.' 'BLAGE'
 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD']
['BILI_T' 'SIXMWT_D' 'SVI' 'ALBUMIN' 'NA.' 'BLAGE'
 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD']
['SIXMWT_D' 'SVI' 'ALK_PHOS' 'BSA' 'SEX_M' 'DIAG_CD_APAH-CVCTD']
```