Principal Investigator

First Name: Rebecca
Last Name: Vanderpool
Degree: PhD, FAHA
Primary Affiliation: The Ohio State University
E-mail: Rebecca.Vanderpool@osumc.edu
State or Province: OH
Country: USA

General Information

Key Personnel (other than PI):
First Name: Raymond
Last name: Benza
Degree: MD
Primary Affiliation: Allegheny Singer Research Institute
SCOPUS ID: 7004568891
Requires Data Access? Unknown

First Name: Manreet
Last name: Kanwar
Degree: MD
Primary Affiliation: Allegheny Singer Research Institute
SCOPUS ID: 23488959300
Requires Data Access? Unknown

First Name: Yongqi
Last name: Liu
Degree: MS
Primary Affiliation: The Ohio State University
SCOPUS ID:
Requires Data Access? Unknown

First Name: Shili
Last name: Lin
Degree: PhD
Primary Affiliation: The Ohio State University
SCOPUS ID:
Requires Data Access? Unknown

How did you learn about the YODA Project?: Colleague

Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2017/07/yoda_project_coi_form_for_data_requestors_2019-_kanwar.pdf
https://yoda.yale.edu/wp-content/uploads/2018/09/coi_form_rv.pdf
https://yoda.yale.edu/wp-content/uploads/2019/12/coi_form_yl.pdf
https://yoda.yale.edu/wp-content/uploads/2019/10/coi_form_rb.pdf
https://yoda.yale.edu/wp-content/uploads/2017/03/coi_form_sl.pdf

Certification

Certification: All information is complete; I (PI) am responsible for the research; data will not be used to support litigious/commercial aims.

Data Use Agreement Training: As the Principal Investigator of this study, I certify that I have completed the YODA Project Data Use Agreement Training

1. NCT01106014 - AC-065A302 - A Multicenter, Double-blind, Placebo-controlled Phase 3 Study Assessing the Safety and Efficacy of Selexipag on Morbidity and Mortality in Patients With Pulmonary Arterial Hypertension
2. NCT00303459 - AC-052-414 (COMPASS-2) - Effects of Combination of Bosentan and Sildenafil Versus Sildenafil Monotherapy on Morbidity and Mortality in Symptomatic Patients With Pulmonary Arterial Hypertension - A Multicenter, Double-blind, Randomized, Placebo-controlled, Parallel Group, Prospective, Event Driven Phase IV Study
3. NCT00433329 - AC-052-419 - COMPASS 3: An Open-label, Multi-Center Study Employing a Targeted 6-Minute Walk Test (6-MWT) Distance Threshold Approach to Guide Bosentan-Based Therapy and to Assess the Utility of Magnetic Resonance Imaging (MRI) on Cardiac Remodeling
4. NCT00091715 - AC-052-364 - A Randomized, Double-blind, Placebo-controlled, Multicenter Study to Assess the Efficacy, Safety, and Tolerability of Bosentan in Patients With Mildly Symptomatic Pulmonary Arterial Hypertension (PAH)
5. NCT00660179 - AC-055-302 - A Multicenter, Double-blind, Randomized, Placebo-controlled, Parallel Group, Event-driven, Phase III Study to Assess the Effects of Macitentan (ACT-064992) on Morbidity and Mortality in Patients With Symptomatic Pulmonary Arterial Hypertension
6. NCT00319267 - AC-052-365 - An Open Label, Multicenter Study to Assess the Pharmacokinetics, Tolerability, and Safety of a Pediatric Formulation of Bosentan in Children With Idiopathic or Familial Pulmonary Arterial Hypertension
7. NCT00319020 - AC-052-367 - An Open Label, Long-term, Safety, and Tolerability Extension Study Using the Pediatric Formulation of Bosentan in the Treatment of Children With Idiopathic or Familial Pulmonary Arterial Hypertension Who Completed FUTURE 1
8. NCT01223352 - AC-052-373 - An Open-label, Prospective Multicenter Study to Assess the Pharmacokinetics, Tolerability, Safety and Efficacy of the Pediatric Formulation of Bosentan Two Versus Three Times a Day in Children With Pulmonary Arterial Hypertension
9. NCT01338415 - AC-052-374 - A Prospective, Multicenter, Open-label Extension of FUTURE 3 to Assess the Safety, Tolerability and Efficacy of the Pediatric Formulation of Bosentan Two Versus Three Times a Day in Children With Pulmonary Arterial Hypertension
10. NCT01743001 - AC-055-305 - A Multi-center, Double-blind, Randomized, Placebo-controlled, Parallel-group, Phase 3 Study to Evaluate the Effects of Macitentan on Exercise Capacity in Subjects With Eisenmenger Syndrome
11. NCT02471183 - AC-065A304 - Multicenter, Open-label, Single-group Study to Assess the Tolerability and the Safety of the Transition From Inhaled Treprostinil to Oral Selexipag in Adult Patients With Pulmonary Arterial Hypertension

Research Proposal

Project Title

Derivation and Validation of Predictive Models in an aggregated Pulmonary Arterial Hypertension cohort

Narrative Summary:

Stratifying risk for pulmonary arterial hypertension (PAH) patients is becoming critical to inform

prognosis and guide treatment choice. Traditional treatment selection relied on clinical gestalt, but the need for an objective risk assessment to guide treatment is supported by expert consensus and research in recent years. Our proposed study will use clinical trial data to derive and validate a novel model for PAH mortality and morbidity predictions. This model will be derived using machine learning methods, including Bayesian modeling, a method that captures the interaction of multiple patient features are they relate to outcomes.

Scientific Abstract:

Background: There is a need for accurate prognostic tools to guide timely and effective therapies for patients with pulmonary arterial hypertension (PAH). Current PAH predictive models are limited due to small derivation populations and traditional modeling methods.
Objective: We will create a harmonized data set from clinical trial data in which to derive and validate predictive models for PAH patient survival, hospitalization, and clinical worsening outcomes.
Study Design: Patient mortality, hospitalization, and clinical worsening outcomes will be evaluated for PAH patients, with the goal of developing predictive models for each outcome. Clinical trial data will be combined into a harmonized dataset, creating a robust source for model derivation and validation.
Participants: This is a study of all Group 1 PAH patients in the requested clinical trials. Inclusion Criteria: Group 1 PAH only, and right heart catheterization diagnostic of pre-capillary pulmonary hypertension.
Main Outcome Measures: Outcomes of interest are patient mortality events, hospitalization events, and clinical worsening events. Models to predict these outcomes will be evaluated by the area under the receiver operating characteristic curve (AUC ROC). The success of stratifying patients by risk outcomes will be demonstrated by Kaplan Meier curves.
Statistical Analysis: Models will be derived using machine learning methods in R. Derived models will be validated on the harmonized data and on datasets outside of Yoda (Registry to Evaluate Early And Long-term PAH Disease Management (REVEAL), etc.)

Brief Project Background and Statement of Project Significance:

Pulmonary Arterial Hypertension (PAH) is a chronic and rapidly progressive disease characterized by extensive narrowing of the pulmonary vasculature leading to progressive increases in pulmonary vascular resistance and eventual death. Approximately half a million persons in the USA will develop PAH over the course of their lifetime. Idiopathic PAH (IPAH) has survival rates at 1, 3, and 5 years of 68%, 48%, and 34%, respectively, with an average survival from onset of symptoms of 2.8 years if left untreated (Hyduk et al. 2005, Sitbon et al. 2002, D?Alonzo et al. 1991)
Even with treatment, the effective 5-year survival is only ~60% among those with PAH enrolled in the Registry to EValuate Early & Long Term PAH Disease Management (REVEAL) (Benza et al. 2012). PAH therapy using initial combination therapies is showing promise in the literature (Ghofrani and Humbert 2014, Galie et al 2009). However, the growing number of options creates a dilemma for determining which treatment is best for a particular patient. This is further complicated by the significant heterogeneity among patients with respect to their clinical responses to available therapies. Therefore, there is a critical need for an accurate, patient-specific prognostic tool to permit tailored, timely, targeted and effective therapies in PAH.
This research proposal is to develop a novel, machine-learning-enabled risk scores for patient risks of mortality, hospitalization, and clinical worsening. This effort builds off the development of the REVEAL and REVEAL 2.0 risk scores (Benza et al, 2010 and Benza et al 2019) by using clinical trial data harmonized to form a larger patient population than currently available in the REVEAL registry. The resulting models will be made available in a web application for use by clinicians to guide patient treatment decision making.

Specific Aims of the Project:

Aim 1: Develop harmonized dataset of clinical trial data and perform feature selection on harmonized data for outcomes of interest (mortality, hospitalization, clinical worsening)

The objective of this Aim is to create a large patient-level dataset for modeling. The hypothesis is that use of a combined dataset will enable inclusion of a larger feature set for modeling and better predictive power from the resulting model, compared to using one trial dataset. Please see Statistical Analyses Plan for further details.

Aim 2: Develop and validate a predictive model for each outcome using machine learning and traditional methods

The objective of this aim is to develop predictive models that perform better than the current risk scores in both internal and external validation, as measured by the area under the receiver operating characteristic curve (AUC ROC). The secondary objective is that the predictive models are able to stratify patients by risk level, as measured by Kaplan-Meier curves. See SAP for more details.

Research Methods

Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:

Data Source: Requested clinical trials: COMPASS 2&3, BENEFIT
This is a study of all Group 1 PAH/ CTEPH patients in the requested clinical trials.
Inclusion Criteria:
? Group 1 PAH
? Right Heart Catheterization (RHC) at enrollment diagnostic of pre-capillary Pulmonary Hypertension (PH) (ie. Mean Pulmonary Arterial Pressure (MPAP) >25mmHg, Pulmonary Capillary Wedge Pressure (PCWP)

Primary and Secondary Outcome Measure(s) and how they will be categorized/defined for your study:

Outcomes of interest:
? Mortality events
? Hospitalization events
? Clinical worsening events, as defined by:
1) Death, Transplantation, Hospitalization due to PH worsening (adjudicated), or Initiation of Prostanoid Therapy/Chronic Oxygen
2) Or, Disease progression defined as all three events occuring:
a. 15% decrease in six minute walk distance (6MWD) from baseline, confirmed by a second walk test on another day
b. A worsening of World Health Organization (WHO) Functional Class (FC) from baseline
c. Addition of a new PAH treatment

Main Predictor/Independent Variable and how it will be categorized/defined for your study:

This assessment will use a combination of independent variables to predict the outcomes listed above (mortality, hospitalization, and clinical worsening) for patients with Group 1 PAH. These independent variables will be from the following categories, as available from each requested clinical trial:
? Laboratory values
? Demographics
? Functional Capacity
? Imaging (Electrocardiogram)
? Hemodynamics and Vitals
? Medical History

Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:

WHO category will be used to group patients for evaluating sub-populations within the Group 1 PAH

cohort.

Patient enrollment in the clinical trial as placebo or intent to treat groups will be used to compare the model performances across groups.

Statistical Analysis Plan:

The clinical trial data will be analyzed by a multi-step process:

? Data will be harmonized into a single dataset, encompassing the requested clinical trials

? Descriptive statistics will be evaluated for all variables

? Variables missingness will be calculated to chose threshold for inclusion and/or imputation

? Variables will be assessed for relationship to outcome of interest using Cox Regression analysis

? Selected variables will be included in predictive modeling for time to outcome using several modeling methodologies in R:

? Random Forest

? Bayesian Network Modeling

? Logistic Regression

? Neural Networks

? Resulting models will be internally evaluated with a 20% withheld test set for performance, as measured by the area under the receiver operating characteristic curve and calibration curve

? Sub-groups in the analysis will be compared by Kaplan-Meier curves to determine significance of risk stratification

? Subgroups may be: WHO category, gender, age group, race, risk scores

? Resulting models will be externally validated on registry data including but not limited to: REVEAL, COMPERA, French PH registry, and the PHSANZ registry

? Resulting models will be externally validated on clinical data from other clinical trials, including by not limited to: AMBITION, FREEDOM-EV, GRIPHON

? Models developed on other PH datasets will be validated in the harmonized dataset

Data harmonization plan:

Data harmonization efforts will focus on unifying key identified features per a previously conducted meta-analysis. Two time points for patients will be considered for harmonization: baseline values and their initial 12-16 week follow-up. Model training will be conducted on two specific endpoints: death (follow-up/last known status) and clinical worsening (end of study). Model training will initially optimize for short term prediction (e.g. one-year survival from baseline and initial follow-up, one-year clinical worsening from baseline and initial follow-up) to maximize the available number of patients with a known status. For early withdrawal/censored patients, status will be imputed as ?alive? or ?event-free?, which has been demonstrated in literature to be a robust imputation strategy provided that early withdrawal patients only compose 20% of the training population. When longer term prediction is desired, inverse proportional censor weighting can be applied to re-create a ?pseudo-population? by replicating patients with long-term known status to effectively replace patients who are censored at an earlier time point, maintaining the same overall population survival curve. This weighting can be determined by modeling the trial itself as a causal effect of early censoring, such that the pseudo-population will then no longer depend on trial follow-up time. Contingency plans will include building temporal models such that early censored patients can still contribute to an accumulation of evidence up until their time of censor.

As part of our ongoing project PHORA, we have developed and validated a risk stratification tool using machine learning to compliment and enhance the traditional methods of analysis (which are typically Cox?s proportional hazard models). We analyzed data of REVEAL 2.0 registry to develop a BN risk model using the variables found in the REVEAL 2.0 calculator with the same discretization cut points at baseline presentation. We then validated it in external registries (e.g. COMPERA) to ensure its validity by creating KM curves separating patients into low, intermediate and high risk based on the 2015 ESC/ ERS guidelines (i.e., low-risk 10% 12-month mortality).

As a next step, we are ?updating? PHORA to allow it to capture additional clinical variables from contemporary PAH clinical trials (such as those being requested from Yoda). As the model learns from a larger number , it will undergo 10-fold cross validation which will be reported as ROC-AUC.

Project Timeline:

Analysis and pre-processing of the data will take place immediately upon receipt of data access

(estimated January 1.) The completion of the harmonization across clinical trials will be completed by the end of January, along with descriptive statistics on the data in the trials. The model derivation will take until the end of December, as multiple modeling modalities will be explored. Once the model is complete, validation will begin on the internal data (February) and external data. Updates to the model may be made throughout this time, with all data analysis and modeling complete by the end of March 2020. Results from analysis will be reported back to YODA by April 30, 2020. The following 3 months will be dedicated to manuscript preparation and writing, as well as abstract submission for CHEST and/or American Heart Association (AHA) conferences. The manuscript submission is expected for June 30th.

Dissemination Plan:

The resulting product will be three predictive models for the three outcomes of interest. These models will be made available for patient risk stratification through the app.myphora.org web platform. The derivation and internal validation of these models will be published in a technical journal and the clinical implications of the models will be published in a clinical journal. The external validation of these models will be published in subsequent paper in a clinical journal.
Target audiences for this work are clinicians, policy makers, and bio-informaticists.
Clinical journals for publication: AHA, CHEST, Circulation
Technical journals for publication: IEEE, AMIA, AIMBE
We also plan to submit segments of the work as abstracts to clinical conferences (AHA, CHEST, PHA, ERS).

Bibliography:

1. Hyduk A, Croft JB, Ayala C, Zheng K, Zheng Z-J, Mensah GA. Pulmonary hypertension surveillance: United states, 1980-2002. US Department of Health and Human Services; 2005.
2. Sitbon O, Humbert M, Nunes H, Parent F, Garcia G, Herv P, Rainisio M, Simonneau G. Long-term intravenous epoprostenol infusion in primary pulmonary hypertension prognostic factors and survival. Journal of the American College of Cardiology. 2002;40:780-788
3. D'Alonzo GE, Barst RJ, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, Fishman AP, Goldring RM, Groves BM, Kernis JT. Survival in patients with primary pulmonary hypertension results from a national prospective registry. Ann Intern Med. 1991;115:343-349
4. Benza RL, Miller DP, Barst RJ, Badesch DB, Frost AE, McGoon MD. An evaluation of long-term survival from time of diagnosis in pulmonary arterial hypertension from the reveal registry survival from time of diagnosis in reveal registry. CHEST Journal. 2012;142:448-456
5. Ghofrani H-A, Humbert M. The role of combination therapy in managing pulmonary arterial hypertension. European Respiratory Review. 2014;23:469-475
6. Gali N, Negro L, Simonneau G. The use of combination therapy in pulmonary arterial hypertension: New developments. European Respiratory Review. 2009;18:148-153
7. Benza RL, Miller DP, Frost A, Barst RJ, Krichman AM, McGoon MD. Analysis of the lung allocation score estimation of risk of death in patients with pulmonary arterial hypertension using data from the reveal registry. Transplantation. 2010;90:298-305
8. Benza RL, Gomberg-Maitland M, Elliott CG, Farber HW, Foreman AJ, Frost AE, McGoon MD, Pasta DJ, Selej M, Burger CD, Frantz RP. Predicting Survival in Patients with Pulmonary Arterial Hypertension. CHEST. Online February 14, 2019