

Principal Investigator

First Name: Sheila
Last Name: McCartan
Degree: PhD, MSc
Primary Affiliation: Exploristics
E-mail: sheila.mccartan@exploristics.com
Phone number: 0044 28 9600 1996
Address: Exploristics, Floor 4, Linenhall Street, Belfast, BT2 8BG
Floor 4, 24 Linenhall Street
City: Belfast
State or Province: N.Ireland
Zip or Postal Code: BT2 8BG
Country: United Kingdom

General Information

Key Personnel (in addition to PI):

First Name: Liam
Last name: Young
Degree: BE(Hons)/BCom
Primary Affiliation: Exploristics
SCOPUS ID:

First Name: Sam
Last name: Miller
Degree: MSc
Primary Affiliation: Exploristics
SCOPUS ID:

First Name: Jamie
Last name: Inshaw
Degree: DPhil, MSc
Primary Affiliation: Exploristics
SCOPUS ID:

First Name: Aisling
Last name: Kearney
Degree: MSc
Primary Affiliation: Exploristics
SCOPUS ID:

First Name: Gareth
Last name: Burns
Degree: PhD
Primary Affiliation: Exploristics
SCOPUS ID:

First Name: Antonia
Last name: Popescu
Degree: MS
Primary Affiliation: Exploristics
SCOPUS ID:

First Name: Jack
Last name: Freebairn
Degree: MS
Primary Affiliation: Exploristics
SCOPUS ID:

Are external grants or funds being used to support this research?: External grants or funds are being used to support this research.

Project Funding Source: Innovate UK

How did you learn about the YODA Project?: Internet Search

Conflict of Interest

https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_ly.pdf
https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_sam_miller.pdf
https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_sheila_mccartan.pdf
https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_jsigned.pdf
https://yoda.yale.edu/system/files/yoda_project_coi_form_for_data_requestors_2019_aisling_kearney_protocol_20204415.docx_2.pdf
https://yoda.yale.edu/system/files/yoda_coi_gb.pdf
https://yoda.yale.edu/system/files/coi_form_ap.pdf
https://yoda.yale.edu/system/files/coi_form_jf.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. [NCT00319111 - AC-052-370 \(BENEFIT OL\) - Long-term Open-label Extension Study in Patients With Inoperable Chronic Thromboembolic Pulmonary Hypertension \(CTEPH\) Who Completed Protocol AC-052-366 \(BENEFIT\)](#)
3. [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](#)
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. [NCT00313222 - AC-052-366 - Prospective, Randomized, Placebo-controlled, Double-blind, Multicenter, Parallel Group Study to Assess the Efficacy, Safety and Tolerability of Bosentan in Patients With Inoperable Chronic Thromboembolic Pulmonary Hypertension \(CTEPH\)](#)
8. [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](#)
9. [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](#)
10. [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](#)
 11. [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](#)
 12. [NCT02070991 - AC-055G201 - A Prospective, Multicenter, Double-blind, Randomized, Placebo-controlled, Parallel-group, 12-week Study to Evaluate the Safety and Tolerability of Macitentan in Subjects With Combined Pre- and Post-capillary Pulmonary Hypertension \(CpcPH\) Due to Left Ventricular Dysfunction](#)
 13. [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](#)

What type of data are you looking for?: Individual Participant-Level Data, which includes Full CSR and all supporting documentation

Research Proposal

Project Title

Development and assessment of Virtual Cohorts derived from historical Pulmonary Arterial Hypertension clinical trials.

Narrative Summary:

Clinical trial simulations carried out in the earliest stages of clinical trial design can significantly reduce development timelines, risks, and costs. Considering the large amount of health care data available there is potential to enhance simulation outcomes by using pre-existing data to generate virtual cohorts. By sourcing data sets associated with pulmonary arterial hypertension (PAH) we propose to optimise and validate a statistical and computational workflow for the creation of a virtual population for use in clinical trial simulation, and in this case, to ultimately inform design of an optimal randomised controlled trial addressing PAH treatment.

Scientific Abstract:

Background: Virtual populations derived from historical clinical trial cohorts can enhance trial simulation leading to better trial design.

Objective: Construct a valid statistical and computational workflow for generating virtual cohorts (VCs), by assessing VC generation from historical pulmonary arterial hypertension (PAH) clinical trial data.

Study Design: VCs generated using a Meta-data model (MM), an Individual model (IM), and a Phenotype model (PM), will be evaluated by investigating heterogeneity for selected characteristics, and through comparison with test cohorts.

Participants: VCs will be constructed from trial baseline participant data from adults; remaining data from cohorts will be used as test data.

Main Outcome Measure(s): Models will be compared by quantifying heterogeneity for VCs, and by comparing against test data cohorts with assumption of no difference between treatment effects.

Statistical Analysis: VC models are generated by Monte Carlo simulation without adjustment (MM), or with adjustment by trial and patient characteristics (IM), or by phenotype (PM). PAH cohort inputs will be streamlined by selection of endpoints (frequency on trial registry), with regression modelling to rank covariate by influence on endpoint. VC heterogeneity is compared through quantification variance parameters before and after application of

each VC model. VC treatment effect estimates of selected endpoints are compared to original cohorts producing both false-positive (control arm comparison) and false-negative error rates (treatment arm comparison).

Brief Project Background and Statement of Project Significance:

In this study, we investigate Pulmonary Arterial Hypertension to help establish a comprehensive statistical and computational pipeline to construct virtual cohorts for employment in clinical trial simulation, with a starting point of individual participant-level data from historical trials.

Simulating a virtual clinical population requires datasets specific to the treatment of interest. This provides patient level information regarding demographics and disease related covariates. These specific datasets can generate a detailed simulation, with more relevance to what is observed in a day-to-day clinical setting, minimising the potential for error and cost in the clinical trial itself.

We will generate virtual cohorts via Monte Carlo simulation using several methodologies including, no adjustment (Meta-data Model), with adjustment by trial and patient characteristics (Individual Model; Ventz et al, 2019), and adjustment by phenotype (Phenotype Model; Klinke, 2008). Each model will be evaluated by investigating VC heterogeneity, and through comparison with positive and negative cohort controls. We hypothesise that VCs derived from the Individual and Phenotype models will more accurately capture the variability inherent within patient populations compared to a meta-data model, while simultaneously reducing the influence of trial to trial bias, providing a more comprehensive approach to clinical trial design.

The results of this study will enhance future clinical trial design, provide an overview of potential statistical and computational approaches to cohort generation, and help to inform decision making criteria with regards to selection of techniques and data sources used to derive virtual cohorts.

Specific Aims of the Project:

Specific Aim: The development and validation of a statistical and computational pipeline for generation of virtual cohorts (VCs) for clinical trial simulation using PAH clinical trial design case study, by means of the following objectives:

1. Generate VCs using three alternative methodologies: Meta-data Model, Individual Model and Phenotype Model.
2. Assess VC design by examining VC heterogeneity and testing for statistically significant differences when compared against positive and negative control cohorts for each model.
3. Compare with actual outcomes of prospective clinical trial.

Hypothesis: VCs derived from the Individual and Phenotype models will more accurately capture the variability inherent within patient populations compared to a meta-data model, proven by less heterogeneity compared to the Meta-data model, and little difference in treatment effect estimates when compared with original trial cohorts.

What is the purpose of the analysis being proposed? Please select all that apply.

Confirm or validate previously conducted research on treatment effectiveness

Participant-level data meta-analysis

Participant-level data meta-analysis using only data from YODA Project

Develop or refine statistical methods

Research on clinical trial methods

Research on comparison group

Research Methods

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

We will include all available trials investigating Pulmonary Arterial Hypertension (PAH).

We would like access to all arms within the trials, developing virtual cohorts using baseline data, including control and treatment arms from Adults with PAH, and then assessing the validity of these virtual cohorts by testing against the remaining trial arms.

Main Outcome Measure and how it will be categorized/defined for your study:

To help refine populations by relevant core characteristics, we will review the clinical trial registry to compile a list of common endpoints and assess their frequency of use across all PAH trials listed, for example, preliminary results show that some of the most investigated endpoints are pulmonary vascular resistance, pulmonary arterial pressure, mortality, cardiac output and 6-minute walking distance. To provide a comprehensive overview of the PAH clinical trial landscape we will investigate each of these endpoints.

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

For each of the endpoints selected by reviewing the clinical trial registry we will run regression modelling to identify covariates of interest.

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

We expect that our initial investigations of frequently used endpoints related to PAH will highlight variables of interest. Relevant patient demographics (age, sex, ethnicity, medical history, etc.) will also be investigated.

Statistical Analysis Plan:

Employing historical data from Randomised Controlled Trials (RCT) and Single-Arm Trials (SAT) in Pulmonary Arterial Hypertension (PAH) we will use three models of virtual cohort (VC) construction to generate VCs for evaluation and use in clinical trial simulation and design: 1) A Meta-data model (MM) where the summary statistics derived across all trials are used to simulate one cohort (MMvc); 2) An Individual Model (IM) where summary data from each cohort are adjusted by trial and patient characteristics (IMvc) before simulation; and finally, 3) A Phenotype Model (PM) where principal component analysis (PCA) is used to define patient clusters derived from patient characteristics, identified as patient phenotype, then summary data from each cohort is adjusted by trial and phenotype (PMvc) before simulation. All models use Monte Carlo simulation to generate VCs, and IM and PM will produce origin matched cohorts which will be pooled after simulation.

First, to help refine populations by relevant core characteristics, we will review the clinical trial registry to compile a list of common endpoints and assess their frequency of use across all PAH trials listed. For commonly used endpoints we will use regression modelling (with method matched to data type, e.g. linear regression for continuous data), to establish covariate influence on key endpoints, ultimately streamlining cohorts by patient characteristics, endpoint and influential covariates.

To give us an indication of variability across cohorts and to assess cohort design, we will investigate and compare heterogeneity of selected characteristics through visualisation of confidence limit overlap via forest plots, and quantification of inconsistency via $I^2 = (Q - df/Q) * 100\%$ with Q the Cochrane Q-statistic, both before and after application of each VC design model.

In each model the VCs are generated from baseline populations in SATs and RCTs (control and treatment arms) in Adults with PAH. The remaining trial arms (post therapy control and treated), as well as paediatric populations from SATs, and an RCT investigating PAH in individuals with Eisenmenger's syndrome, will be used as positive controls, i.e. we will expect differences when comparing these populations to VCs at particular endpoints due to influence of treatment effects, time points or patient characteristics. The VCs will also be tested by comparison with VC subsets grouped by the original trial in a 'a leave one out' approach, this will be considered as a negative control, i.e. we should find no differences when comparing these populations to the VCs. In all cases comparisons will be made by randomly selecting sample sizes determined by a targeted power of 80% and 10% Type I error rate and estimating the treatment effect by adjusting with inverse variance weighting and testing null hypothesis of no-benefit at a targeted type I error rate of 10%. There will be 10,000 iterations of each step with this sample size. The outputs will be used to inform clinical design of a prospective PAH trial, and ultimately, we aim to compare all results to the actual trial outcome.

Software Used:

RStudio

Project Timeline:

Project can begin when initial data sets are received.

Preliminary data analysis: 4 months (investigating key covariates, implementing VC models to build VCs ,and investigating heterogeneity).

Testing VCs and trial simulation: 3 months.

Interpretation of results, production of draft manuscript: 2 months.

Publication ready for submission: 1 month.

Total project time requirement: 10 months.

Dissemination Plan:

It is anticipated that there is scope to publishing findings in relation to two areas of this project:

Virtual cohort design and Clinical trial simulation; and Specific outcomes from the RCT of PAH treatment.

With regards to our findings on clinical trial simulation and design we hope to publish in journals such as CPT: Pharmacometrics & Systems Pharmacology, BMC Medical Informatics and Decision Making and Trials, for example, which have precedent for publishing on clinical trial design. We can also consider general journals such as PLOS one or F1000. Components of the pipeline can also be published as standalone methods and protocols. We aim to highlight key findings by poster or presentation at relevant conferences such as PSI, Bio.IT World Conference, Health Systems Research (HSR), and Intelligent Health, for example.

Scientific outcomes specifically relating to the design and implementation of the PAH clinical trial will be published in a highly ranked, peer-reviewed international biomedical journal.

We aim to pursue an open access publication strategy.

Bibliography:

Klinke (2008). Integrating Epidemiological Data into a Mechanistic Model of Type 2 Diabetes: Validating the Prevalence of Virtual Patients. *Annals of Biomedical Engineering*, 36: 321-334. doi: 10.1007/s10439-007-9410-y

Ventz S, Lai A, Cloughesy TF, Wen PY, Trippa L & Alexander BM (2019). Design and evaluation of an external control arm using prior clinical trials and real-world data. *Clin Cancer Res*, 25(16): 4993 – 5001. doi: 10.1158/1078-0432.CCR-19-0820.