Principal Investigator

First Name: Abhinav
Last Name: Sharma
Degree: MD, PhD
Primary Affiliation: McGill University
E-mail: amir.razaghizad@mail.mcgill.ca
Phone number: 5145737940
Address: 5252 de Maisonneuve Boulevard West

City: Montreal
State or Province: Quebec
Zip or Postal Code: H4A 3S9
Country: Canada

General Information

Key Personnel (in addition to PI):
First Name: Amir
Last name: Razaghizad
Degree: MSc
Primary Affiliation: McGill University
SCOPUS ID:

First Name: Jiayi
Last name: Ni
Degree: MA / MS / MSc
Primary Affiliation: Research Institute of McGill University Health Center
SCOPUS ID:

First Name: Pedro
Last name: Marques
Degree: MD
Primary Affiliation: McGill University Health Center
SCOPUS ID:

First Name: Thomas
Last name: Mavrakanas
Degree: MD, MSc, FRCPC, FASN
Primary Affiliation: McGill University
SCOPUS ID:

First Name: Frederic
Last name: Baroz
Degree: MD
Primary Affiliation: McGill MUHC
SCOPUS ID:

Are external grants or funds being used to support this research?: No external grants or funds are being used to support this research.
How did you learn about the YODA Project?: Colleague

Conflict of Interest

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. NCT01032629 - A Randomized, Multicenter, Double-Blind, Parallel, Placebo-Controlled Study of the Effects of JNJ-28431754 on Cardiovascular Outcomes in Adult Subjects With Type 2 Diabetes Mellitus
2. NCT01989754 - A Randomized, Multicenter, Double-Blind, Parallel, Placebo-Controlled Study of the Effects of Canagliflozin on Renal Endpoints in Adult Subjects With Type 2 Diabetes Mellitus
3. NCT02065791 - A Randomized, Double-blind, Event-driven, Placebo-controlled, Multicenter Study of the Effects of Canagliflozin on Renal and Cardiovascular Outcomes in Subjects With Type 2 Diabetes Mellitus and Diabetic Nephropathy

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

Research Proposal

Project Title

Cluster Analysis of Cardiovascular Phenotypes and SGLT2 Inhibition in Patients With Type 2 Diabetes and Established Cardiovascular Disease

Narrative Summary:

Diabetes is the 9th leading cause of death worldwide. Although it has been studied for several decades, it remains a strong risk factor for illnesses affecting the heart and kidneys, even when treated. Evidence suggests that this may be because several unknown types (or ?compositions?) of the disease exist. As a result, our study proposes using machine learning to uncover hidden compositions of the illness, which may have different prognoses and reactions to treatment. These findings will help doctors provide personalized care by identifying patients who may need increased antidiabetic care, and likewise, patients who may benefit most from certain treatments.

Scientific Abstract:

Background In the CANVAS and CREDENCE trials,8,9 canagliflozin was associated with 33 and 39% reduction in the incidence of heart failure hospitalization, respectively. However, whether differential treatment effects exist with canagliflozin remains unclear.
Objective As a result, this study aims to utilize race and latent class analysis to identify distinct clinical phenotypes in subjects with T2D and cardiovascular disease to elucidate potential differences

in treatment effects across race and clinical phenotypes.

Study Design Latent class analysis will be utilized to identify unobserved (or ?latent?) subclasses of individuals with T2D and cardiovascular disease. Cox proportional hazard regression models with interaction terms will be utilized to assess whether cluster membership is associated with a differential response to canagliflozin. Subgroup analyses according to race will also be done.

Participants The population of interest for the proposed analysis encompasses the entire patient populations enrolled in the CANVAS and CREDENCE trials. In analyses where differential survival according to cluster membership is evaluated, patients will be subdivided into their respective trial arms (i.e., canagliflozin, placebo).

Main Outcomes The co-primary endpoints for the proposed analysis are time to heart failure hospitalization, and the composite of heart failure hospitalization or cardiovascular death. Secondary endpoints include cardiovascular death, nonfatal myocardial infarction, nonfatal stroke, cause specific mortality, all-cause mortality, and death from renal causes.

Statistical Analysis We will assess proportional hazards assumptions, and survival models will be adjusted for baseline clinical characteristics.

Brief Project Background and Statement of Project Significance:

Despite significant advances in the design of therapies for people with type 2 diabetes mellitus (T2DM), the disease continues to portend high rates of morbidity and mortality, even when traditional cardiovascular risk factors are well controlled.[2] Like disease states such as heart failure and atherosclerosis, there is significant evidence that this may be because numerous pathophysiological phenotypes of the disease exist.[3] Although diabetes has historically been diagnosed into two classes (i.e., type I and type II), these data suggest that our binary approach to treatment may not be sufficient for risk reduction. As a result, unsupervised learning (e.g., machine learning) algorithms such as latent class analyses have proliferated in the clinical literature to improve personalized care.

As an alternative to traditional subgroup analysis, latent class analyses are a data agnostic method for elucidating distinct clinical phenotypes and their associated response to treatment.[4] This method has been utilized successfully in a variety of medical disciplines, including heart failure with reduced or preserved ejection fraction.[5,6] More recently, cluster analysis was utilized retrospectively in the Empaglifozin, Cardiovascular Outcome, and Mortality in T2DM (EMPA-REG outcomes) trial, identifying a nonsignificant trend of greater benefit in one of three phenotypes as characterized by lower rates of heart failure hospitalization in young people with low comorbidity burden.[7] Despite these findings, some uncertainty persists as relatively few patients experienced these endpoints.

Thus, we propose conducting a pooled latent class analysis in patients from the CANVAS (Canagliflozin Cardiovascular Assessment Study)[8] and CREDENCE (Canagliflozin and Renal Events in Diabetes with Established Nephropathy Clinical Evaluation)[9] trials to assess the differential effects of SGLT2 inhibition on cardiovascular and renal outcomes in patients with T2DM and established cardiovascular disease. Such an analysis could aid in the identification of phenotypes (or ?clusters?) of people with T2DM who may respond most beneficially to SGLT2 inhibition. This could help researchers identify novel patient groups for investigating new glycemic and cardiovascular treatments, potentially saving valuable resources, and allowing health care providers to offer greater access to care.

Specific Aims of the Project:

Primary Aim:
1. To evaluate if distinct phenotypes of patients with type 2 diabetes can be identified in the CANVAS and CREDENCE trials, through hierarchical latent class analysis.
2. To evaluate if the race of patients with type 2 diabetes delineates a unique clinical phenotype with distinct differential outcomes and treatment responses.
Secondary Aims:
1. To evaluate if these clusters have differential outcomes with regards to CANVAS and CREDENCE?s endpoints (i.e., composite of cardiovascular death, nonfatal myocardial infarction, or nonfatal stroke; heart failure hospitalization; cardiovascular death or heart failure hospitalization; cardiovascular death; all-cause mortality; and death from renal causes)?
2. To evaluate if these clusters have differential treatment responses to canagliflozin with regards to

CANVAS and CREDENCE?s endpoints (i.e., composite of cardiovascular death, nonfatal myocardial infarction, or nonfatal stroke; heart failure hospitalization; cardiovascular death or heart failure hospitalization; cardiovascular death; all-cause mortality; and death from renal causes).
Tertiary Aims:
1. To evaluate

Research Methods

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

For primary aim #1, which assesses whether distinct phenotypes exist, all patients enrolled in the CANVAS and CREDENCE trials will be eligible, and clusters will be stratified according to randomization. Race stratification will also be done for primary aim #2.
For secondary aim #1, which assesses whether clusters have differential outcomes, all patients enrolled in the CANVAS and CREDENCE trials will be eligible, and clusters will be stratified according to randomization.
For secondary aim #2, which assesses whether clusters have differential treatment responses, only the patients enrolled in CANVAS and CREDENCE?s canagliflozin arm will be evaluated.

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

The co-primary endpoint of interest for our analysis will be time to heart failure hospitalization, and the composite of cardiovascular death or heart failure hospitalization as defined, respectively, in CANVAS and CREDENCE.
The secondary endpoints of interest include nonfatal myocardial infarction, or nonfatal stroke; heart failure hospitalization; cardiovascular death; all-cause mortality; and death from renal causes.

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

The main predictor/independent variable for all analyses will be the clusters identified with the latent class analysis algorithm. As defined in the ?Statistical Analysis Plan? the number of clusters will be selected based on clinical significance (i.e., whether the clusters are clinically distinct in terms of baseline characteristics), the Bayesian information criterion, and the size of the smallest cluster. These criteria has been empirically shown to yield the best results.15

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

Common variables/risk factors available across the CANVAS and CREDENCE trials will be utilized for the latent class analysis as well any regression models evaluating differential prognosis and treatment response. For the latent class analysis, all baseline variables/risk factors will be put through an oblique principal component analysis (PROC VARCLUS procedure in SAS)12 to normalize and reduce the dimensionality of the data to meet the presuppositions of latent class analysis.13 That is, the SAS procedure will aggregate all available variables into several non-overlapping clusters, which are defined by a summary score characterized by a linear combination of the variables from each patient. (The coefficients of the variable cluster summary score are identified by the first principal component of the variable cluster, which is to be done separately for continuous and categorical variables.) For Cox proportional hazards regression models, the variables of interest (i.e., baseline clinical, vital, and laboratory characteristics) will be unaltered and defined per the criteria in CANVAS and CREDENCE. Where non-overlapping definitions are observed, clinical experts will be consulted.

Statistical Analysis Plan:

Clinical Variable Selection and Data Cleaning

Prior to beginning the analysis, baseline clinical variables will be jointly examined by two study investigators (AS, AR) to assess overlap between CANVAS and CREDENCE. In addition, we will evaluate and remove variables that are highly collinear and/or have limited clinical availability or relevance. We will also remove any variables that are coded as positive in less than 10% of cases as these have been shown to negatively affect patient clustering.[13] Following initial variable assessment, multiple imputation (n=5) with the Markov chain Monte Carlo method will be performed if variables have a moderate proportion of missing data (e.g., 40% of data are missing) or if the ?missing not at random? assumption is plausible, only complete cases will be used.[16]

Data Preparation for Latent Class Analysis

Following clinical variable selection and data cleaning, dimension reduction will be performed on the covariate list. This will be done through independent oblique principal component analysis with the SAS PROC VARCLUS procedure.[12,17] In brief, the PROC VARCLUS procedure is an iterative variable clustering process that continuously divides a set of variables into disjoint clusters. That is, ?clusters that are as correlated as possible among themselves and as uncorrelated as possible with variables in other clusters?.[12,17] This process will be applied to categorical and continuous variables independently. To determine the appropriate number of variable clusters, we will iteratively evaluate the second eigenvalue of each covariate group. A stopping rule (eigenvalue threshold of 0.7) will be used as suggested by Jackson and colleagues.[18] Each patient?s covariate cluster will ultimately be defined by a normalized principal component summary score (i.e., a linear combination of variables) which will be used in the latent class analysis. Analogously, each patient will be defined by a matrix of covariate summary scores allowing for their clustering.

Patient Clustering and Latent Class Analysis

Following data preparation, we will subsequently identify the latent clusters of individuals. This will be done utilizing the ?poLCA? package in R, which utilizes a latent class analysis algorithm.[14] In brief, latent class analysis is a statistical model that classifies individuals into mutually exclusive (and exhaustive) clusters based on their observed set of measured characteristics. These clusters will be derived using a maximum likelihood estimation. To avoid finding a local maximum of the log-likelihood function, the model will be estimated 10 times to automate the search for the global maximum. To derive the optimal number of clusters or subgroups, we will evaluate the first minima of the Bayesian information criteria, the size of the smallest class, and the clinical relevance of defined groups.[15] We will use an a priori criteria of at least 200 patients per cluster to promote stability of effect estimates.

Inferential Statistics

We will first assess the association between cluster membership and clinical outcomes using Cox proportional hazards regression. Proportional hazards assumptions will be assessed by evaluating scaled Schoenfeld residuals. Kaplan-Meier estimated mortality and cause-specific event rates will be plotted according to treatment allocation. Using interaction terms in a Cox regression model, we will also evaluate whether cluster membership is associated with a differential response to randomized therapy. Results will be reported for clinical outcomes as hazard ratios (HR) and 95% confidence intervals (CI) for each cluster in comparison to a reference cluster and an overall p-value is provided to assess the relationship between cluster and outcome. In models with interactions, an interaction p-value will be provided. P

Project Timeline:

We anticipate completing all study milestones within approximately 8 months, as shown in detail in the Gantt chart (supplemental material). If provided the data, we would start the project on January 15th, 2022. All analyses, including consultations with cardiologists and nephrologists, would be completed in approximately the first 12 weeks. Then, we anticipate spending 4 weeks to write the initial draft of the manuscript and spending another 4 weeks inviting co-authors to help edit the text. The estimated time of completion of the finalized manuscript would thus be June 15th, 2022, at the latest. We would then develop a journal submission plan, which would include general and specialty medical journals including, but not limited, to the Journal of the American Medical Association (JAMA), the European Heart Journal (EHJ), Circulation, and the American Heart Journal. During this time, we would contemporaneously prepare abstracts for major conferences like the American College of Cardiology (ACC) and the Global Cardiovascular Clinical Trialists Forum (CVCT). Once the manuscript is accepted for publication, we would report our results to back to the YODA project by August 15th,

2022, at the latest.

Dissemination Plan:

Our knowledge translation and dissemination plan is a critical aspect of our proposal as it will ensure that results impact health service delivery, clinical care, and future development of novel therapies. As described above, we plan to write manuscript(s) for peer-reviewed medical journals, which we will widely distribute through academic social media and national or international conferences. We would target general and specialty medical journals including, but not limited, to the Journal of the American Medical Association (JAMA), the European Heart Journal (EHJ), Circulation, the Canadian Journal of Cardiology, and the American Heart Journal. During this time, we would contemporaneously prepare abstracts for major scientific conferences such as the American College of Cardiology (ACC) Annual Scientific Session & Expo, and the Global Cardiovascular Clinical Trialists Forum (CVCT).

Bibliography:

1. World Health Organization. The top 10 causes of death. Accessed June 23, 2021. https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death
2. Rawshani A, Rawshani A, Franzen S, et al. Risk Factors, Mortality, and Cardiovascular Outcomes in Patients with Type 2 Diabetes. New England Journal of Medicine. 2018;379(7):633-644. doi:10.1056/NEJMoa1800256
3. Gouda P, Zheng S, Peters T, et al. Clinical Phenotypes in Patients With Type 2 Diabetes Mellitus: Characteristics, Cardiovascular Outcomes and Treatment Strategies. Curr Heart Fail Rep. 2021;18(5):253-263. doi:10.1007/s11897-021-00527-w
4. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. Statistics in medicine. 1986;5(1):21-27.
5. Ahmad T, Pencina MJ, Schulte PJ, et al. Clinical implications of chronic heart failure phenotypes defined by cluster analysis. J Am Coll Cardiol. 2014;64(17):1765-1774. doi:10.1016/j.jacc.2014.07.979
6. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA. Latent Class Analysis of ARDS Subphenotypes: Analysis of Data From Two Randomized Controlled Trials. Lancet Respir Med. 2014;2(8):611-620. doi:10.1016/S2213-2600(14)70097-9
7. Sharma A, Ofstad AP, Ahmad T, et al. Patient Phenotypes and SGLT-2 Inhibition in Type 2 Diabetes. JACC: Heart Failure. 2021;9(8):568-577. doi:10.1016/j.jchf.2021.03.003
8. Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. New England Journal of Medicine. 2017;377(7):644-657.
9. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and Renal Outcomes in Type 2 Diabetes and Nephropathy. New England Journal of Medicine. 2019;380(24):2295-2306. doi:10.1056/NEJMoa1811744
10. Hagenaars JA, McCutcheon AL. Applied Latent Class Analysis. Cambridge University Press; 2002. Accessed October 8, 2021. http://ebookcentral.proquest.com/lib/mcgill/detail.action?docID=217833
11. Linda M. Collins, Stephanie T. Lanza. Latent Class and Latent Transition Analysis. 1st ed. John Wiley & Sons, Ltd; 2009. doi:10.1002/9780470567333
12. SAS Help Center: Overview: VARCLUS Procedure. Accessed October 12, 2021. https://documentation.sas.com/doc/en/pgmsascdc/9.4_3.3/statug/statug_var...
13. Sinha P, Calfee CS, Delucchi KL. Practitioner?s Guide to Latent Class Analysis: Methodological Considerations and Common Pitfalls. Crit Care Med. 2021;49(1):e63-e79. doi:10.1097/CCM.0000000000004710
14. Linzer DA, Lewis JB. poLCA: An R package for polytomous variable latent class analysis. Journal of statistical software. 2011;42(1):1-29.
15. Nylund KL, Asparouhov T, Muthn BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling: A Multidisciplinary Journal. 2007;14(4):535-569. doi:10.1080/10705510701575396
16. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials ? a practical guide with flowcharts. BMC Med Res Methodol. 2017;17(1):1-10. doi:10.1186/s12874-017-0442-1
17. Nelson BD. Variable Reduction for Modeling using PROC VARCLUS. Data Analysis.:3.

18. Jackson JE, Edward A. User?s guide to principal components. John Willey Sons. Inc, New York. Published online 1991:40.

19. Clark SL, Muthn B. Relating Latent Class Analysis Results to Variables Not Included in the Analysis.; 2009.