Principal Investigator

First Name: Mansu
Last Name: Kim
Degree: Ph.D.
Primary Affiliation: Gwangju Institution of Science and Technology
E-mail: mansu.kim@gist.ac.kr
State or Province: Gwangju-gwangyeoksi
Country: Republic of Korea

General Information

Key Personnel (other than PI):
First Name: Jae-Seung
Last name: Yun
Degree: M.D. Ph.D.
Primary Affiliation: The Catholic University of Korea
SCOPUS ID:
Requires Data Access? Yes

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?: Scientific Publication

Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2024/07/240710-COI-mskim-1.pdf
https://yoda.yale.edu/wp-content/uploads/2024/07/240711-COI-YJS-1.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. NCT01032629 - 28431754DIA3008 - 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 - 28431754DIA4003 - 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

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

Research Proposal

Project Title

Trans-ethnic effect of Canagliflozin Cardiovascular risk in Type 2 Diabetes based on phenomapping derived too

Narrative Summary:

We aim to investigate the effect of Canagliflozin on cardiovascular risk across ethnic groups (i.e., East Asian and Caucasian) in Type 2 Diabetes. A previous study developed a predictive tool for drug response using phenomapping from the CANVAS trial. Specifically, the collected key variables, including categorical and continuous, are embedded in a shared latent space to harmonize and calculate distances between individuals. We will then use a phenomapping-derived tool, a powerful tool for predicting individualized risk, to measure individualized hazard ratios within neighborhoods. Finally, we compare individualized risks to examine canagliflozin on cardiovascular risk across ethinic.

Scientific Abstract:

Background: SGLT2 inhibitors like canagliflozin provide cardioprotective benefits in type 2 diabetes but are underused due to high costs and the lack of individualized treatment strategies. The differential effects of canagliflozin across ethnic groups, particularly East Asian and Caucasian populations, remain underexplored.
Objective: To investigate the trans-ethnic effects of canagliflozin on cardiovascular risk in type 2 diabetes, using phenomapping to compare individualized risk predictions across East Asian (Korean) and Caucasian populations.
Study design: A retrospective cohort study comparing CANVAS trial data with a matched Korean EHR dataset. Phenomapping will be applied independently to both datasets to identify differences in canagliflozin response across ethnicities.
Participants: The study includes 10,135 participants from the CANVAS and CANAVAS-R trials and 11,100 from the Korean EHR dataset, comprising canagliflozin users and matched non-users.
Primary and Secondary Outcome Measure(s): Primary outcome: The primary outcome is the time to first major adverse cardiovascular event (MACE), including cardiovascular death, nonfatal myocardial infarction, and nonfatal stroke. Secondary outcomes include hospitalization for heart failure and progression of renal disease.
Statistical analysis: A topological representation of the study population will be constructed using baseline variables and used for caluclating similarity of inidividuals. Cox regression models and SHAP values will assess hazard ratios and feature importance across ethnic groups.

Brief Project Background and Statement of Project Significance:

- Project Background: Type 2 diabetes is a global health crisis, leading to a significantly higher risk of CVD, the leading cause of death worldwide. SGLT2 inhibitors, such as canagliflozin, have shown substantial cardioprotective effects in patients with type 2 diabetes [1]. Despite their benefits, these medications are underutilized, often due to high costs and the lack of tailored treatment strategies that consider diverse responses across different ethnic groups.
Ethnic differences in drug responses are a critical area of research, particularly in the context of type 2 diabetes. The CANVAS trial, a key study on the cardiovascular effects of canagliflozin, primarily included Caucasian participants (77.8%), with only 13.4% Asian representation. This disparity raises concerns about the applicability of the trial’s findings to other ethnic groups, especially East Asians. Understanding whether the cardiovascular benefits of canagliflozin observed in primarily Caucasian populations hold true in East Asian populations is crucial for optimizing treatment strategies.
This research aims to address this gap by investigating the trans-ethnic effects of canagliflozin on cardiovascular risk in type 2 diabetes among East Asian (Korean) and Caucasian populations. We will use a 10-year follow-up dataset of 125,000 Korean patients, matched using propensity scores to create a real-world evidence (RWE) cohort comparable to the CANVAS trial. Although our data is not from a randomized controlled trial (RCT), the matched cohort allows for a robust comparative

analysis between the two ethnic groups.
- Project Significance: This project is significant for its potential to advance generalizable scientific and medical knowledge, particularly regarding the differential impact of canagliflozin across ethnic groups. By applying the same phenomapping methodology independently to the CANVAS trial and the matched Korean dataset, we will generate comparative insights into whether the cardiovascular benefits of canagliflozin are consistent across racial groups or if significant differences exist [2].
Our phenomapping tool will integrate key clinical variables into a shared latent space [3,4], enabling the calculation of distances between individuals and the prediction of individualized cardiovascular risk. This approach will help identify specific patient phenotypes that benefit most from canagliflozin, leading to more personalized and effective treatment strategies.
The insights gained from this research will contribute to the broader goal of precision medicine, where treatment decisions are better tailored to individual patient profiles, thereby maximizing therapeutic benefits and optimizing resource use. Additionally, these findings could inform public health strategies, ensuring more equitable and effective healthcare policies that address the unique needs of diverse populations.
In summary, this project will significantly enhance our understanding of the ethnic-specific effects of canagliflozin on cardiovascular risk in type 2 diabetes, contributing to improved clinical outcomes and advancing the field of precision medicine.

Specific Aims of the Project:

Our objective is to apply the same phenomapping methodology independently to the two datasets (the CANVAS trial and the matched Korean EHR dataset) and then compare the results. We aim to determine whether the response to SGLT2 inhibitors is consistent across different racial groups or if there are significant differences. To clarify, we do not intend to merge the two cohorts for analysis; instead, we will analyze them separately using the same methodology, presenting two distinct results for comparison.

Study Design:

Individual trial analysis

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

Confirm or validate previously conducted research on treatment effectiveness

Research on clinical prediction or risk prediction

Research Methods

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

For the first dataset (i.e., CANVAS trials), we plan to collect comprehensive baseline variables and follow-up data for patients with type 2 diabetes. The inclusion criteria for the CANVAS trial are as follows: 1) Age: 30 years or older with established ASCVD or 50 years or older with two or more ASCVD risk factors; 2) Diagnosis: HbA1c levels between 7.0% and 10.5% (53 to 91 mmol/mol) in patients diagnosed with type 2 diabetes; 3) ASCVD risk factors: Presence of two or more of the following risk factors. The exclusion criteria included type 1 diabetes, severe disease, history of UTI or GTI, DKA, hypersensitivity to canagliflozin or its components, pregnancy, lactation, and low life expectancy. For the second dataset, we have collected a hospital-based type 2 diabetes cohort established at three general hospitals in South Korea. In this dataset, variables such as age, gender, smoking status, BMI, blood pressure, dyslipidemia, and glycated hemoglobin will be extracted from canagliflozin users and non-users using 1:2 propensity matching. The same exclusion criteria as those used in the CANVAS trial will be applied to this dataset for analysis and comparison.

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

your study:

The primary outcome measure for this study is the time to the first occurrence of a composite of major adverse cardiovascular events. This composite includes Cardiovascular death, nonfatal myocardial infarction, and Nonfatal stroke. Secondary outcomes include hospitalization for heart failure and progression of renal disease.

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

Variables collected by both institutions are considered in this study.
- Demographic and Clinical Features: Age, Sex
- Medical conditions: Dyslipidemia, Hypertension, Retinopathy, Nephropathy, Neuropathy, Coronary artery disease, Cerebrovascular disease, Peripheral arterial disease, History of amputation, History of coronary revascularization
- Medication Usage: Diuretics, Beta blockers, Renin-angiotensin-aldosterone blockers, Calcium channel blockers, Antiplatelets, Anticoagulants, Statins Insulin (any type, intermediate-long acting, short-acting), Biguanides, Thiazolidinediones, Dipeptidyl-peptidase IV inhibitors, Glucagon-like peptide 1 receptor agonists, Other antihyperglycemic agents
- Lifestyle and Physical Measurements: Smoking history, Weight, Pulse, Systolic and Diastolic blood pressure, Height, (BMI)
- Blood and Urine Tests: Various biomarkers including Albumin, Alkaline phosphatase, Alanine aminotransferase, Aspartate aminotransferase, Basophil count and percentage, Bicarbonate, Bilirubin, BUN, Calcium, HDL, LDL, Creatine kinase, Chloride, C-peptide, Creatinine (serum and urine), Eosinophil count and percentage, GFR, GGT, Glucose, HbA1c, Insuline, Potassium, and etc.

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

None

Statistical Analysis Plan:

First, we plan to propose a modified Gower’s method to find the phenotypic similarity of all individuals. Specifically, we will use the weighted Gower method, which computes a metric of dissimilarity between two data points, including both numerical (e.g., age, BMI) and non-numerical data (e.g., history of hypertension, smoking status). The weighting in the Gower method will be adjusted to give appropriate importance to different types of variables based on their clinical relevance and statistical properties.

Second, once the phenotypic similarities are calculated, individuals will be grouped into phenotypically similar clusters. These clusters will be defined as neighborhoods encompassing 5% to 30% of the most similar individuals to each index individual. Within these neighborhoods, we will conduct multivariate Cox regression models to estimate individualized hazard ratios (HR) for major adverse cardiovascular events (MACE). The resulting HRs will be mapped onto a high-dimensional space and visualized using t-SNE (t-distributed Stochastic Neighbor Embedding). This visualization will allow us to compare the HR maps between East Asian and Caucasian populations to identify any significant differences in risk patterns.

Third, we will develop an individualized prediction model using a penalized regression approach, such as LASSO (Least Absolute Shrinkage and Selection Operator) or elastic net regression. This model will be trained to predict cardiovascular risk based on the phenotypic data. The model’s performance will be rigorously evaluated using nested five-fold cross-validation to ensure robustness and generalizability. To investigate the trans-ethnic effect of canagliflozin, we will apply the individualized risk prediction model developed from the Caucasian population to the East Asian population and vice versa. This cross-application will help us understand how well the risk factors and treatment effects generalize across different ethnic groups.

For model interpretability, we will use SHAP (Shapley Additive Explanations) values to assess feature importance. SHAP values provide a unified measure of the contribution of each feature to the prediction, making it easier to interpret the influence of individual variables on the predicted cardiovascular risk. This will help us identify which phenotypic characteristics are most critical in determining the individualized risk and how they differ between ethnic groups.

By following this detailed and rigorous approach, we aim to uncover important insights into the trans-ethnic effects of canagliflozin on cardiovascular risk in type 2 diabetes, ultimately contributing to more personalized and effective treatment strategies.

Software Used:

Python

Project Timeline:

In our study, data cleaning and preprocessing will be conducted over three months. Based on the phenomapping-derived tool, the model will be developed, trained, and tested over the following four months. Interpretation of results and sensitivity analyses will be conducted, and manuscript drafting will occur in the subsequent two months, with the first submission planned for nine months after the project begins. Extensions will be requested if additional time is needed.

Dissemination Plan:

We plan to prepare and submit multiple manuscripts detailing the study's findings. The primary manuscript will focus on the overall trans-ethnic effects of canagliflozin on cardiovascular risk in type 2 diabetes. Additional manuscripts may explore specific aspects of the data, such as subgroup analyses and detailed methodological approaches. Potentially Suitable Journals are Diabetes Care, The Lancet Diabetes & Endocrinology, Journal of the American Medical Association (JAMA), The New England Journal of Medicine (NEJM), Circulation, and European Heart Journal.

Bibliography:

[1] Neal B, Perkovic V, Mahaffey KW, de Zeeuw D, Fulcher G, Erondu N, Shaw W, Law G, Desai M, Matthews DR. Canagliflozin and Cardiovascular and Renal Events in Type 2 Diabetes. N Engl J Med 2017;377:644-57.

[2] Oikonomou EK, Suchard MA, McGuire DK, Khera R. Phenomapping-Derived Tool to Individualize the Effect of Canagliflozin on Cardiovascular Risk in Type 2 Diabetes. Diabetes Care 2022;45:965-74.

[3] Gower JC. A general coefficient of similarity and some of its properties. Biometrics 1971;27:857–871

[4] McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. Accessed 28 January 2022. Available from https://arxiv.org/abs/1802.03426

Supplementary Material:

https://yoda.yale.edu/wp-content/uploads/2024/08/Response-Letter.pdf