Principal Investigator

First Name: Linong
Last Name: Ji
Degree: MD
Primary Affiliation: Peking University People's Hospital
E-mail: jiln@bjmu.edu.cn
State or Province: Beijing
Country: China

General Information

Key Personnel (other than PI):
First Name: Xiantong
Last name: Zou
Degree: MD PhD
Primary Affiliation: Peking University People's Hospital
SCOPUS ID: 55964605900
Requires Data Access? Yes

First Name: Qi
Last name: Huang
Degree: MBCHB
Primary Affiliation: Peking University People's Hospital
SCOPUS ID:
Requires Data Access? Yes

First Name: Zhouhui
Last name: Lian
Degree: PhD
Primary Affiliation: Peking University
SCOPUS ID:
Requires Data Access? No

First Name: Sarah
Last name: Wild
Degree: PhD
Primary Affiliation: The University of Edinburgh
SCOPUS ID:
Requires Data Access? No

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: Beijing Nova Cross program (Z211100002121169); The National Natural Science Foundation of China (T2341011)
How did you learn about the YODA Project?: Scientific Publication

Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2024/05/JLN_disclosure.pdf
https://yoda.yale.edu/wp-content/uploads/2024/05/ZXT_disclosure.pdf
https://yoda.yale.edu/wp-content/uploads/2024/05/HQ_disclosure.pdf

https://yoda.yale.edu/wp-content/uploads/2024/05/LZH_disclosure.pdf
https://yoda.yale.edu/wp-content/uploads/2024/05/sarah_disclosure.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. NCT02065791 - 28431754DNE3001 - 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
2. 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
3. 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

Identify treatment responders in patients with type 2 diabetes using a machine learning based dynamic cardiovascular risk assessment tool ML-CVD in clinical trials

Narrative Summary:

Type 2 diabetes is a common health concern characterised by high blood sugar levels. It affects over 500 million adults worldwide and is associated with a significant escalation of the cardio-renal complications. With the development of novel antihyperglycemic medications such as SGLT2 inhibitors, GLP-1 receptor agonists, and MRAs (Mineralocorticoid Receptor Antagonists), more targeted and efficacious treatments have been utilized. Now clinical guidelines recommend special tools to calculate the probability of developing cardiorenal events. These tools are designed to categorize patients into low, medium, or high risk, and the initiation of novel protective drugs is preferred for patients at high risk. However, these predictive tools weren’t always accurate enough for people with type 2 diabetes, especially for those already have and at high risk for cardiorenal disease.
In our previous study, we created a new tool called ML-CVD. This tool used patient information and health updates over time to have a more precise prediction on the 5-year probability of cardiovascular disease than the traditional tools. It also helped to identify which patient benefit more from SGLT2 inhibitor. But whether this model could predict the treatment effect of other drugs is unknown.
In this study, we aim to validate the ML-CVD model in both controlled clinical trials and real-world settings. We want to know if it's effective to predict cardiovascular risk for a broader patient group and if it can also identify those who might have more benefit from specific medications. The study could lead to better care for people with type 2 diabetes by providing more accurate risk assessments and identifying more beneficial treatments for each individual.

Scientific Abstract:

Background: Recent evidence suggested that glucagon-like peptide-1 receptor agonists (GLP-1 RA), sodium-glucose cotransporter 2 inhibitors (SGLT2i), and nonsteroidal mineralocorticoid receptor antagonists (ns-MRA) were options significantly reducing cardio-renal risks for patients with type 2 diabetes. Traditional cardiovascular disease (CVD) risk tools provided a strategy for identifying high-risk patients but they lacked the ability to make accurate prediction and estimate individual benefits from protective medications. Our prior research developed a machine learning (ML) model for CVD prediction that surpassed conventional tools in forecasting major cardiovascular events and heart failure and identify treatment responders. The model needs for further validation in broader cardiovascular outcome trials.
Objectives: (1) To validate the discrimination of the ML-CVD model on predicting cardiovascular risk; (2) To identify the responder of SGLT2i, GLP-1 RA, dipeptidyl peptidase-4 inhibitor (DPP4i), metformin and ns-MRA; (3) To identify the cutoff value of ML-CVD to predict positive drug effect on outcomes.
Study Design: In each cardiovascular outcome trial, we would include the intent-to-treat population. As we would use baseline factors and interim risk factors post-intervention, we defined a 1-year observational window post-intervention for longitudinal data collection. Individuals who experienced cardiovascular events or who were censored in the first year were excluded.
Participants: Patients taking medication for at least 52 weeks and having complete information on outcome would be included.
Main Outcome Measure: 4p-MACE (nonfatal stroke, nonfatal myocardial infarction, cardiovascular death, and hospitalization for heart failure), renal outcome (doubling of serum creatinine, 40% reduction in eGFR, or end-stage kidney disease).
Statistical Analysis: For validation, we would assess Harrell’s C-index and the area under the receiver operating characteristic curve (AUROC) at half-year intervals during the follow-up period to assess the discriminative ability of the model. We also compared the performance of the ML-CVD model with that of traditional risk scores. We would establish an ML-CVD score to predict the 5-year probability of the cardiovascular events.To assess the patients’ response to specific treatments, the ML-CVD scores were calculated at baseline and 1 year later. Patients were categorised into 'responders' and 'non-responders' based on the change in their ML-CVD scores over the 1-year observation period: A decrease in the ML-CVD score indicated a 'responder', and 'non-responders' experienced no decrease. We would subsequently analyse the actual composite cardiovascular events at the end of the trial between both groups using HRs and 95% CIs from the Cox regression models, adjusting for the initial risk levels. To analysis the association between metformin and cardiovascular outcome, patients prescribed metformin either at baseline or at any subsequent study visit would were classified as the ‘metformin use’ group, whereas those who never received metformin were classified as the 'non-use' group. We would use multivariable Cox model, inverse probability of treatment weighting (IPTW) and targeted maximum likelihood estimation (TMLE) to estimate the association of metformin use and the outcomes. To understand the implications of ML-CVD score differences between the treatment arm and placebo arm on the RCT estimates of the treatment effects, we would involve simulations and determine cutoff value for the ML-CVD score to predict protective effect for specific medication.

Brief Project Background and Statement of Project Significance:

The management of cardio-renal complications is a major challenge in type 2 diabetes treatment. Metformin, long recognized for its glucose-lowering properties, were considered as potential beneficial for preventing long-term cardiovascular events. Recent randomized controlled trials have shown glucagon-like peptide-1 receptor agonists (GLP-1 RA), sodium-glucose cotransporter 2 inhibitors (SGLT2i), and nonsteroidal mineralocorticoid receptor antagonists (ns-MRA) are more effective options for reducing cardio-renal risks1–3; DPP4i has confirmed cardiovascular safety. In current practice, it was recommended to use cardiovascular disease (CVD) risk tools, such as Framingham risk score, pooled cohort equation (PCE) and SCORE-2D for stratification and initiate protective medications in high-risk patients4,5. Yet, these tools mainly derived from the general population, exhibiting unsatisfactory discriminative capability when applied to diabetes population and failing to account for individual benefits from protective medications6.
In our previous study, we found that a model incorporating baseline risk factors and interim changes significantly enhanced cardiovascular prediction accuracy. In 2022, we applied a proposal on Vivli (ID

00007844), utilised data from canagliflozin clinical trials and ACCORD to create a machine learning (ML) model for cardiovascular disease (CVD) prediction. We proposed using data from canagliflozin clinical trials to develop a machine learning (ML) model for CVD prediction on Vivli. This model surpassed traditional scores in precisely predicting major cardiovascular incidents and heart failure, while also facilitating ongoing risk monitoring throughout the intervention period. By tracking the risk change estimated by ML-CVD, we could identify responders to intensive glycemic therapy and canagliflozin who exhibited greater risk reduction under intervention than the non-responders. Despite these promising results, further validation in other cardiovascular outcome trials and real-world cohorts is necessary.
This study aims to validate the performance of ML-CVD model and its ability on identification of treatment responders across cardiovascular outcome trials (CVOTs) of SGLT2i, GLP-1 RA, dipeptidyl peptidase-4 inhibitor (DPP4i), and ns-MRA. We will compare the model's external discrimination with that of traditional tools and assess the ML-CVD score (5-year probability of cardiovascular outcome) respectively at baseline and after one year of treatment. We will evaluate the change of ML-CVD scores across different interventions, encompassing medications that protect against cardiovascular issues like SGLT2i, GLP-1 RA, and ns-MRA, as well as medications known for their cardiovascular safety, such as DPP4i, and the potentially protective medication, metformin. The ML-CVD score for patients receiving these treatments will be calculated at baseline and after the 1-year observation period. Those with a decreased score will be categorized as "responders," while those whose score remains unchanged or increases will be considered as "non-responders." We will then analyze the incidence rates of cardiovascular events among these groups to determine if responders, as identified by the ML-CVD score, derive greater benefits from specific treatments compared to non-responders. Evaluating the impact of drugs on cardiovascular outcomes demands significant labor, time, and resources. We aim to assess the potential of the ML-CVD score as a sensitive indicator of a beneficial cardiovascular preventive effect. By examining association between ML-CVD score and the actual cardio-protective effect of various drugs, we will be able to identify the ML-CVD score's cutoff value for predicting cardio-protective effects.

Specific Aims of the Project:

We hypothesize that patients have various risks for cardiovascular outcome and responses to specific medications. By employing the ML-CVD model, we aim to predict the risk and identify subgroups that may derive greater benefits from specific medications.
Aim:
(1) To validate the discrimination of the ML-CVD model on predicting cardiovascular risk in different cardiovascular outcomes trials and real-world evidence.
(2) To identify the responder of SGLT2i, GLP-1 RA, DPP4i, metformin and ns-MRA;
(3) to identify the cutoff value of ML-CVD to predict positive drug effect on cardiorenal outcomes.

Study Design:

Individual trial analysis

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

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:

To validate our machine learning models, we will use cardiovascular outcome trials (CVOTs) of SGLT2i, GLP-1 RA, DPP4i, and ns-MRA. We have submitted our data-request on Vivli platform and also put in data request on the YODA platform as data on CANVAS, CANVAS-R and CREDENCE were owned by Johnson & Johnson. If both organizations approved our request, our analysis will be performed on Vivli platform (https://vivli.org/) .

All subjects in the intent-to-treat (ITT) analysis set were included in our analyses. Main exclusion criteria included：(1) patietnts experienced cardiovascular events within the first year. (2) patients censored within the first year.

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

Cardiovascular diseases are major long-term complications in people with type 2 diabetes. We have developed a machine-learning model ML-CVD to predict the cardiovascular outcome in patients at high risk. In this study, we would like to validate the model so we focus on the specific endpoints of type 2 diabetes. The primary outcome included the composite of cardiovascular death, nonfatal myocardial infarction, nonfatal stroke, or hospitalisation for heart failure. Secondary outcomes included:
(1) the first occurrence of cardiovascular death.
(2) the first occurrence of nonfatal myocardial infarction.
(3) the first occurrence of nonfatal stroke.
(4) the first occurrence of hospitalisation for heart failure.
(5) if available, we would analysis the renal composite: doubling of serum creatinine, 40% reduction in eGFR, or end-stage kidney disease.
All endpoints were adjudicated by specialised committees across the trials. Patients who did not experience the primary outcome at the end of the trial would be censored.

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

We extracted the baseline clinical features related to cardiovascular from the trials, both categorical and continuous, as shown in Table 1. Data included demographics, physical measurements, disease history, and laboratory measurements. During each patient visit, we identified the interim risk factors inferred to mediate the effects of anti-glycaemic therapies on the cardiovascular outcomes in multiple clinical trials7–9. These encompassed physical measurements, such as body mass index (BMI), systolic blood pressure (SBP), and diastolic blood pressure (DBP), as well as laboratory assessments including haemoglobin A1C (HbA1c), estimated glomerular filtration rate, serum creatinine, urine albumin-creatinine ratio (UACR), triglycerides, total cholesterol, high-density lipoprotein cholesterol (HDL), and low-density lipoprotein cholesterol (LDL), which were identified as cardiovascular risk determinants or therapeutic biomarkers for the effectiveness of SGLT2i. Furthermore, intervention details including the timing of follow-up visits quantified as a continuous variable and the binary classification of canagliflozin usage ('Yes' or 'No') were also included.
Table 1 Variable for validation
1. Pretreatment risk factors
Demographics features age, sex, race (categorized as White, Asian, Black, Other), smoking status (categorized as Yes/No)
Physical measurements height, weight, BMI, SBP, DBP, heart rate
Disease history (categorized as Yes/No for each) hypertension, heart failure, hyperlipidemia, cardiovascular disease, microvascular disease (including chronic kidney disease, diabetic eye disease, neuropathy)
Medication history (categorized as Yes/No for each) insulin, other anti-diabetic drugs, anti-hypertensive, anti-lipidemia, other cardiovascular drugs
Laboratory measurements HbA1c, eGFR, SCr, ALT, UACR, TG, TC, HDL, LDL
2. Interim risk factors at visits
Physical measurements BMI, SBP, DBP
Laboratory measurements HbA1c, eGFR, SCr, ALT, UACR, TG, TC, HDL, LDL
Intervention after baseline Taking protective medication (Yes/No), follow-up time

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:

(1) Data preparation
We defined a 1-year observational window post-intervention for longitudinal data collection. Individuals who experienced cardiovascular events or who were censored in the first year were excluded. For missing data at the baseline, we utilised multiple imputation strategy. We didn't process missing data at follow-up in the external testing dataset. Non-normalised variables were log-transformed, followed by normalisation of all the numeric variables to ensure comparability prior to the analysis.
(2) Model validation
We would assess Harrell's C-index and the area under the receiver operating characteristic curve (AUROC) at half-year intervals during the follow-up period to assess the discriminative ability of the model. We also compared the performance of the ML-CVD model with that of traditional risk scores, such as the Framingham risk score, PCE and SCORE-2D. We would establish an ML-CVD score to predict the 5-year probability of the cardiovascular events during the follow-up period.
(3) Dynamic surveillance of the cardiovascular risks
We would calculate the mean ML-CVD scores and their 95% CIs for each treatment group in each trial. The differences in the ML-CVD scores across the groups at 12 months were assessed using the Student's t-test adjusted for the ML-CVD scores at baseline.
(4) Identify treatment responders
We would assess the patients' response to SGLT2i, GLP-1 RA, DPP4i and ns-MRA. The ML-CVD scores were calculated at baseline and 1 year later to ascertain the change in the patient's ML-CVD score. The association between the change in the ML-CVD score and cardiovascular outcomes was estimated using the Cox regression models. Patients were categorised into 'responders' and 'non-responders' based on the change in their ML-CVD scores over the 1-year observation period. A decrease in the ML-CVD score indicated a 'responder', implying reduced cardiovascular risk, and 'non-responders' experienced no decrease, implying stable or increased risk. We would subsequently analyse the actual composite cardiovascular events at the end of the trial between both groups using HRs and 95% CIs from the Cox regression models, adjusting for the initial risk levels.
Metformin held the most potential benefit for the prevention of CV events. Per designed of each trial, patients prescribed metformin either at baseline or at any subsequent study visit would were classified as the 'metformin use' group, whereas those who never received metformin were classified as the 'non-use' group. All the pretreatment risk factors described in above section would be considered as covariates for the metformin use and endpoints, and we would use multivariable Cox model, inverse probability of treatment weighting (IPTW) and targeted maximum likelihood estimation (TMLE) to estimate the association of metformin use and the outcomes. We also calculated the change of ML-CVD within the 1-year observation window and separated patients in the 'metformin use' group into 'responder' and 'non-responder' to analyse the actual composite cardiovascular events at the end of the trial using Cox regression models.
(5) Identify the cutoff value of protective effect
We intended to pool data separately for each category of medication, including SGLT2i, GLP-1 RA, DPP4i, and ns-MRA, from their respective CVOTs. Our approach would involve simulations in which ML-CVD score differences were adjusted using random sampling with replacement. Fluctuations in these differences will likely alter the span of the 95% confidence interval for the hazard ratio and reflect the change of treatment effectiveness compared with placebo. We would determine the exact ML-CVD score difference that resulted in the upper limit of the hazard ratio dropping below 1 (suggesting a protective effect from the treatment). This cutoff value for the ML-CVD score will be instrumental in predicting a beneficial drug impact on clinical outcomes for a specific category of medication.

Software Used:

RStudio

Project Timeline:

0-9 month: finish model validation
9-12 month: analysis the association between metformin and cardiovascular outcome

Dissemination Plan:

We plan to publish a manuscriptina high impact peer-reviewed journal e.g. Lancet diabetes and endocrinology or Diabetes Care.
We also plan to submit our abstract toaninternational conference on diabetes including ADA or EASD.

Bibliography:

1. 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. doi:10.1056/NEJMoa1611925
2. Marso SP, Daniels GH, Brown-Frandsen K, et al. Liraglutide and cardiovascular outcomes in type 2 diabetes. *New England Journal of Medicine*. 2016;375(4):311-322.
3. Bakris GL, Agarwal R, Anker SD, et al. Effect of finerenone on chronic kidney disease outcomes in type 2 diabetes. *New England Journal of Medicine*. 2020;383(23):2219-2229.
4. American Diabetes Association Professional Practice Committee. 10. Cardiovascular Disease and Risk Management: Standards of Medical Care in Diabetes—2022. *Diabetes Care*. 2021;45(Supplement_1):S144-S174. doi:10.2337/dc22-S010
5. Marx N, Federici M, Schütt K, et al. 2023 ESC Guidelines for the management of cardiovascular disease in patients with diabetes. *European Heart Journal*. Published online August 25, 2023:ehad192. doi:10.1093/eurheartj/ehad192
6. Ndumele CE, Rangaswami J, Chow SL, et al. Cardiovascular-Kidney-Metabolic Health: A Presidential Advisory From the American Heart Association. *Circulation*. 2023;148(20):1606-1635. doi:10.1161/CIR.0000000000001184
7. Buse JB, Bain SC, Mann JFE, et al. Cardiovascular risk reduction with liraglutide: An exploratory mediation analysis of the leader trial. *Diabetes Care*. 2020;43(7):1546-1552. doi:10.2337/dc19-2251
8. Inzucchi SE, Zinman B, Fitchett D, et al. How does empagliflozin reduce cardiovascular mortality? Insights from a mediation analysis of the EMPA- REG OUTCOME trial. *Diabetes Care*. 2018;41(2):356-363. doi:10.2337/dc17-1096
9. Agarwal R, Tu W, Farjat AE, et al. Impact of finerenone-induced albuminuria reduction on chronic kidney disease outcomes in type 2 diabetes: a mediation analysis. *Annals of internal medicine*. 2023;176(12):1606-1616.

Supplementary Material:

https://yoda.yale.edu/wp-content/uploads/2024/05/Graphic-Abstract.docx