

Principal Investigator

First Name: Linong

Last Name: Ji

Degree: M.D.

Primary Affiliation: Peking University People's Hospital

E-mail: Jiln@bjmu.edu.cn

Phone number:

Address: No. 11 Xi Zhi Men Nan Da Jie, West District, Beijing 100044

City: Beijing

State or Province: Beijing

Zip or Postal Code: 100044

Country: China

General Information

Key Personnel (in addition to PI):

First Name: Xiantong

Last name: Zou

Degree: MD PhD

Primary Affiliation: Peking University People's Hospital

SCOPUS ID: 55964605900

First Name: Qi

Last name: Huang

Degree: MD candidate

Primary Affiliation: Peking University People's Hospital

First Name: Zhouhui

Last name: Lian

Degree: PhD

Primary Affiliation: Peking University

First Name: Edward

Last name: Boyko

Degree: MD MPH

Primary Affiliation: University of Washington

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) to Xiantong Zou: Precision medicine in diabetes facilitated by artificial intelligence.

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

Conflict of Interest

https://yoda.yale.edu/system/files/hq_disclosure_0.pdf

https://yoda.yale.edu/system/files/zxt_disclosure_0.pdf

https://yoda.yale.edu/system/files/jln_disclosure_0.pdf

https://yoda.yale.edu/system/files/eb_disclosure.pdf

https://yoda.yale.edu/system/files/lzh_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. [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](#)
3. [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](#)

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

Research Proposal

Project Title

Risk stratification and responder identification for GLP-1RA and SGLT2i in T2DM: application of two machine learning based prediction models

Narrative Summary:

Kidney failure, heart attacks and stroke are the main causes of premature death in diabetes patients. Recently, two drugs, named GLP-1 RA and SGLT2i, showed robust effects in changing the kidney and heart complications of diabetes. But not all diabetes patients develop those complications, so which person should choose GLP-1RA and which person should choose SGLT2i were unknown. We aim to develop artificial intelligence facilitated tools to help clinicians identify patients who were prone to cardio-renal complications and who responder better to SGLT2i or GLP-1RA. Accordingly, we will help clinicians find the most fitting patients and make a good choice on drugs.

Scientific Abstract:

Background: Glucagon-like peptide-1 receptor agonists (GLP-1 RA) and sodium-glucose cotransporter 2 inhibitors (SGLT2i) were drugs to alter cardio-renal outcomes, however, there was a growing clinical demand to predict patients' response to these drugs as early as possible. Conventional cardiovascular risk prediction models were mainly developed using baseline features, but the cardiovascular risk altered over time in a dynamic way.

Objectives: To develop a baseline risk model and a dynamic risk model, which incorporated additional interim risk factors, to stratify patients' cardio-renal risks and predict their responses to SGLT2i and GLP-1RA treatment.

Participants: Patients taking medication for at least 52 weeks and having complete information on outcome would be included.

Study Design: We will develop two machine learning models in a large cardiovascular outcome trial (CVOT) ACCORD and validate this model in albiglutide and canagliflozin CVOT and renal outcome trials (ROT) (Harmony, CANVAS, CANVAS-R and CREDENCE). The baseline risk model will use only baseline clinical features and the dynamic risk model will incorporate additionally interim features in early treatment phase. Various machine learning methods including Xgboost and ANN will be applied.

Main Outcome Measure: 3p-MACE, renal outcome compositions (40% reduction in eGFR, renal-replacement therapy, or renal death), and progression of albuminuria.

Statistical Analysis: The accuracy will be assessed by area under the curve (ROC) and C-statistic in ACCORD. Models with the best performance will be applied in the validation cohorts. The baseline risk model will be used for risk stratification in all patients in Harmony and CANVAS, and the heterogeneity of treatment-by-group will be tested in patients within subgroups stratified by their predicted progression risk. Patients in the active drug-treated arm who were predicted to encounter an event by the baseline risk model but were absent from any event during the whole follow-up period were defined as 'eventual responders'. Patients predicted to have a subsequent reduction in the event risk during the 1st year of active drug treatment predicted by the dynamic risk model will be defined as 'interim responders'. The concordance rate and characteristics between eventual responders and interim responders of the two drugs will be compared.

Brief Project Background and Statement of Project Significance:

Cardiovascular disease and diabetic nephropathy are major long-term complications in people with T2DM. In the last decade, various clinical trials suggested that glucagon-like peptide-1 receptor agonists (GLP-1 RA) and sodium-glucose cotransporter 2 inhibitors (SGLT2i) prevented the progression of cardio-renal endpoints versus placebo [1,2,3]. These drugs were recommended as first-line therapy for T2DM patients with high cardiovascular risks and a tendency to nephropathy by many societies, e.g. European Society of Cardiology (ESC) [4,5]. However, in the setting of clinical trials, nearly all recruited participants were at high risk for cardiovascular disease [1-3]. Only 15-20% of patients of high-risk develop a cardiovascular event or renal event within 3-5 years. These patients also showed differential responses to GLP-1RA /SGLT2i therapy. Identifying the risk of cardiovascular disease and diabetes nephropathy and choosing relative therapy on individual level have been an important part of personalized medication for T2DM.

Previous models for predicting cardio-renal events in participants with T2DM focused on patients' baseline characteristics [6]. These models used pretreatment variables in other clinical trials, including the demographics data, physical examination, and blood and urine tests, showing high performance on prediction [7, 8]. But disease progression is a highly heterogeneous and dynamic process [9], so outcome prediction based on an assessment at a fixed time point may lead to misestimation in some patients. To improve the predictive performance, dynamic predictive models using subsequent follow-up data may be imposed to guide clinical choice in many diseases such as lymphoma [10]. Recent studies have identified that SGLT2/GLP1-RA treatments result in favourable effects on cardiovascular risk factors, including blood glucose, blood pressure, weight and albuminuria [11], which may be associated with a reduction of the risks of cardiovascular complications and kidney diseases. More secondary analysis suggested these factors may mediate the cardio-renal protective effect of these two drugs, so they were also candidate interim risk factors [12,13]. Adding these interim risk factors in the early cycles of treatment, especially factors potentially mediating cardio-renal benefit of drugs, will improve the prediction of long-term outcomes and estimation of drug-specific benefits.

We hypothesize that alleviation in interim risk factors within a short treatment period is a strong prognostic factor for long-term response in terms of cardio-renal event prevention. By adding this information to machine learning models, our study helps to precisely predict a patient's cardio-renal outcomes and their responses to drugs.

Specific Aims of the Project:

- (1) Develop and validate a baseline risk prediction model and dynamic risk prediction model for cardio-renal outcomes in CVOTs.
- (2) To stratify patients with T2DM into different cardio-renal risk categories and compare their responses to GLP-1RA/SGLT2i.
- (3) To identify participants who may respond to GLP-1RA or SGLT2i therapy at an early phase of treatment.

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

New research question to examine treatment effectiveness on secondary endpoints and/or within subgroup populations

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 develop our machine learning models, we will use ACCORD (NCT00000620) as the derivation cohort and Harmony (NCT02465515), CANVAS (NCT01032629), CANVAS-R (NCT01989754) and CREDENCE (NCT02065791) as external validation cohorts. We have submitted our data-request on Vivli platform for ACCORD and HARMONY and CANVAS, CANVAS-R and CREDENCE. We 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 resuest, 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 inclusion criteria included?(1) taking medication for at least 52 weeks. (2) complete information on cardiovascular and renal outcome (defined in the next section).

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

In this study, we would analyze the cardiovascular outcome trials of T2DM to (1) stratify patients with different outcome progression risks and (2) try to identify drug responders. The primary outcome of our study was 3p-MACE (yes/no): a composite of nonfatal stroke, nonfatal myocardial infarction, and cardiovascular death. Secondary outcomes included: (1) Renal outcome (yes/no): a composite of 40% reduction in eGFR, renal-replacement therapy, or renal death. (2) Progression of albuminuria (yes/no): 30% increase in UACR and change from either normoalbuminuria to microalbuminuria, or macroalbuminuria, or from microalbuminuria to macroalbuminuria. The event rate at the end of trials will be estimated by the logistic regression model and the 5-year probability of outcome will be estimated by our machine-learning methods.

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

We will develop two models to predict the cardio-renal progression risks. In baseline risk model, variables will be selected from the demographics data (such as age, sex, and country), physical examination (such as body weight, height, BMI, waist circumference, pulse, SBP, DBP), and blood and urine test (such as fasting plasma glucose, HbA1c, fasting insulin, fasting c-peptide, TG, TC, HDL, LDL, hs-CRP, UACR, uric acid), physical activity, smoking status, family history of cardiovascular disease, and other medication.

Our dynamic risk model considers predictors during the treatment, including pretreatment risk factors and interim risk factors at early follow-up cycles (less than 1 year). Interim risk factors will be selected from potential contributors to final cardio-renal prevention as described in previous articles [12, 13]. The interim risk factors may include body weight, BMI, confirmed hypoglycemia, SBP, HbA1c, TG, TC, HDL, LDL, eGFR, UACR, uric acid and other anti-diabetes drug use.

Statistical Analysis Plan:

(1) Data preparation and variable selection

In each cohort, continuous variables will be standardized and skew variables will be log-transformed. Missing data will be generated by multiple imputations. Variables with missing data in more than 20% of samples or high collinearity (Pearson coefficient>0.7) will be first excluded.

(2) Development and Validation of baseline risk models

We will develop baseline risk models with pretreatment variables in ACCORD trial. It was shown in ACCORD that intensive glycemic therapy didn't reduce the risk for 3p-MACE or composite renal outcomes, except for the progression of albuminuria [14], so we combine the intensive glycemic control arm and standard glycemic control arm for cardiovascular disease and renal composite outcomes. However, intensive glycemic therapy reduced the risk for the progression of albuminuria [15]. We will add the treatment method as a variable to deviate the model for the albuminuria progression endpoint.

Some machine learning methods will be applied, such as random forest, k-nearest neighbours (kNN), support vector machine (SVM), naïve Bayes, Elastic Net, XGBoost and if appropriate, artificial neural network (ANN). Models will be calibrated and validated using 10-fold cross-validation. Model performance will be assessed by area under the curve (ROC) and C-statistic and the model with the best performance will be selected as the final "baseline risk model". The cut-off point of this model will be determined using the Youden index. This model will be externally tested in the placebo arm of all validation cohorts since the outcome may be altered by SGLT2i/GLP-1RA treatment.

(3) Application 1: Cardio-renal Risk stratification within T2DM

We will apply our model in validation cohorts, i.e. HARMONY and CANVAS, to separate all participants into high-risk and low-risk groups using optimal cut-off points. Treatment effect heterogeneity will be tested between the high risk and low-risk groups on the absolute scale by estimating absolute risk differences using logistic regression, and

on the relative scale by comparing models including an interaction term with models excluding interaction term using likelihood-ratio tests.

(4) Development and Validation of dynamic risk models

We will add interim risk factors into the same baseline risk model to estimate updated outcome risks during different disease courses. The refined new model will be named “Dynamic Risk Model” and externally tested on all participants in validation cohorts, regardless of glucose-lowering strategies. To assess the advantage of our new model, we will also compare the performance of our models with our previously established baseline risk model, and with other well-established cardiovascular risk scores, such as SCORE (Systematic Coronary Risk Evaluation) and atherosclerotic cardiovascular disease algorithm for 10-year risk based on Pooled Cohort Equations (ASCVD) [8,9] in our validation populations.

(5) Application 2: Drug Responder identification

The biggest challenge for clinicians was to identify specific patients responding to SGLT2i or GLP-1RA at the beginning. To help to solve it, we will incorporate the information at different stages of treatment into our dynamic risk models. Gathering patients’ information at baseline, 1st follow-up and 2nd follow-up, our model will estimate the patients’ risks for outcome 3 times (recorded as risk0, risk1 and risk2) in a dynamic manner. A patient estimated to have a decreased risk during disease courses (i.e. 30% risk reduction from baseline) will be defined as an ‘interim responder’. We also define actually final “responder” as a patient who was predicted to have a cardio-renal event at baseline but failed to occur at the end of the trial in GLP-1RA or SGLT2i treated arm. The concordance rate will be measured between final responders and interim responders and the characteristics of interim responders will be presented as an exploratory analysis.

Software Used:

RStudio

Project Timeline:

2022.6~2024.6

1st year: finish model development using ACCORD, CANVAS, CANVASR, CREDENCE and HARMONY.

2nd year: finish model application and submit manuscript date

Dissemination Plan:

We plan to publish a manuscript in a high impact peer-reviewed journal e.g. Lancet diabetes and endocrinology or diabetes care.

We also plan to submit our abstract to an international 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. *N Engl J Med.* 2017;377(7):644-657. doi:10.1056/NEJMoa1611925.
2. Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and Renal Outcomes in Type 2 Diabetes and Nephropathy. *N Engl J Med.* 2019;380(24):2295-2306. doi:10.1056/NEJMoa181174.
3. Hernandez AF, Green JB, Janmohamed S, et al. Albiglutide and cardiovascular outcomes in patients with type 2 diabetes and cardiovascular disease (Harmony Outcomes): a double-blind, randomised placebo-controlled trial. *Lancet.* 2018;392(10157):1519-1529. doi:10.1016/S0140-6736(18)32261-X
4. Frank L J Visseren, et al. 2021 ESC Guidelines on cardiovascular disease prevention in clinical practice. *Eur Heart J.* 2021 Aug30;ehab484. doi: 10.1093/eurheartj/ehab484.
5. Cosentino F, Grant PJ, Aboyans V, Bailey CJ, Ceriello A, Delgado V, et al. 2019 ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur Heart J* 2020;41:255-323. doi: 10.1093/eurheartj/ehz486.
6. Rossello X, Dorresteijn JA, Janssen A, et al. Risk prediction tools in cardiovascular disease prevention: A report from the ESC Prevention of CVD Programme led by the European Association of Preventive Cardiology (EAPC) in collaboration with the Acute Cardiovascular Care Association (ACCA) and the Association of Cardiovascular Nursing and Allied Professions (ACNAP). *Eur Heart J Acute Cardiovasc Care.* 2020;9(5):522-532. doi:10.1177/2048872619858285
7. Piepoli MF, Hoes AW, Agewall S, et al. 2016 European guidelines on cardiovascular disease prevention in clinical practice. The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of G Ital Cardiol (Rome) 2017;18:547–612.
8. Budoff MJ, Young R, Burke G, et al. Ten-year association of coronary artery calcium with atherosclerotic cardiovascular disease (ASCVD) events: the Multi-Ethnic Study of Atherosclerosis (MESA). *Eur Heart J*

2018;39:2401–8.

9. Li FR, Yang HL, Zhou R, et al. Diabetes duration and glycaemic control as predictors of cardiovascular disease and mortality. *Diabetes Obes Metab*. 2021;23(6):1361-1370. doi:10.1111/dom.14348

10. Kurtz DM, Esfahani MS, Scherer F, et al. Dynamic Risk Profiling Using Serial Tumor Biomarkers for Personalized Outcome Prediction. *Cell*. 2019;178(3):699-713.e19. doi:10.1016/j.cell.2019.06.011

11. Vasilakou D, Karagiannis T, Atha-nasiadou E, et al. Sodium-glucose co-transporter 2 inhibitors for type 2 diabetes: a systematic review and meta-analysis. *Ann Intern Med* 2013;159:262-74.

12. 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;41:356–63. doi:10.2337/dc17-1096

13. Buse JB, Bain SC, Mann JFE, et al. Cardiovascular risk reduction with liraglutide: An exploratory mediation analysis of the leader trial. *Diabetes Care* 2020;43:1546–52. doi:10.2337/dc19-2251

14. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med*. 2008;358(24):2545-2559.

doi:10.1056/NEJMoa0802743

15. Mottl AK, Buse JB, Ismail-Beigi F, et al. Long-Term Effects of Intensive Glycemic and Blood Pressure Control and Fenofibrate Use on Kidney Outcomes. *Clin J Am Soc Nephrol*. 2018;13(11):1693-1702.

doi:10.2215/CJN.06200518