Use of glycaemic and lipid variability to predict the risk for major adverse cardiovascular events in patients with type 2 diabetes mellitus

Dr Sharen Lee

Registration Number: 100193659

In submission for the award of PhD by Publication

Norwich Medical School, University of East Anglia

May 18th, 2024

Primary Supervisor: Professor Vassilios Vassiliou

Secondary Supervisor: Associate Professor Johannes Reinhold

Word Count: 43 442

Table of contents

Acknowledgements 6
Abstract 7
List of tables 9
List of figures 11
List of abbreviations 12
Chapter 1. Introduction 14
1.1. Background 14
1.2. The role of glycaemic control in cardiovascular risk control 15
1.3. The role of lipid control in cardiovascular risk control 18
1.4. Role and contributors to temporal variability 20
1.5. Conclusion 22
Chapter 2. Predictions of diabetes complications and mortality using Hba1c variability: a 10-year observational cohort study 23
2.1. Introduction 23
2.2. Methods 24
2.2.1. Study population 24
2.2.2. Dataset description 24
2.2.3. Patient data 25
2.2.4. Statistical analysis 27
2.3. Results 28
2.3.1. Clinical and biochemical characteristics 28
2.3.2. HbA1c variability and baseline 30
2.4. Discussion and limitations 36
2.4.1. Discussion 36
2.4.2. Limitations 39
2.5. Conclusion 40
Chapter 3. Glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning 41
3.1. Introduction 41
3.2. Methods 42
3.2.1. Study population 42
3.2.2. Patient data 42
3.2.3. Statistical analysis 43
3.2.4. Development of a regularized and weighted random survival forest model 44

3.3. Results 46
3.3.1. Clinical and biochemical characteristics 46
3.3.2. Anti-diabetic drug classes and outcomes 50
3.3.3. Adverse outcome and predictors 51
3.3.4. The relationship between neutrophil-lymphocyte ratio, frequency of hypoglycaemic episodes and glycaemic variability 67
3.3.5. Survival learning results 68
3.4. Discussion and limitations 81
3.4.1 Discussion 81
3.4.2. Strengths and limitations 86
3.5. Conclusion 87
Chapter 4. Development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong 88
4.1. Introduction 88
4.2. Methods 89
4.2.1. Study design and data source 89
4.2.2. Data extraction 89
4.2.3. Variability calculations 90
4.2.3. Outcomes and statistical analysis 91
4.2.4. Development of machine/ deep models for survival learning 92
4.3. Results 94
4.3.1. Baseline characteristics 94
4.3.2. Predictors of all-cause mortality 97
4.3.3. Development of a score-based predictive risk model based on Cox regression 106
4.3.4. Results of machine/ deep learning approaches for risk modelling 108
4.4. Discussion and limitations 111
4.4.1. Discussion 111
4.4.2. Strengths and limitations 115
4.5. Conclusion 116
Chapter 5. Risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 117
5.1. Introduction 117
5.2. Methods 118
5.2.1. Ethics approval and method overview 118
5.2.2. Patient data 118
5.2.3. Statistical analysis 120

5.3 Results 122
5.3.1. Baseline characteristics 122
5.3.2. Predictors of sudden cardiac death and atrial fibrillation 126
5.4. Discussion and limitations 133
5.4.1. Discussion 133
5.4.2. Strengths and limitations 135
5.5. Conclusion 136
Chapter 6. Predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death 137
6.1. Introduction 137
6.2. Methods 138
6.2.1. Study design 138
6.2.2. Data extraction 138
6.2.3. Statistical analysis 139
6.3. Results 141
6.3.1. Baseline characteristics 141
6.3.2. Acute myocardial infarction prediction 144
6.3.3. Sudden cardiac death prediction 150
6.3.4. Machine learning survival analysis 156
6.4. Discussion and limitations 160
6.4.1. Discussion 160
6.4.2. Limitations 163
6.5. Conclusion 164
Chapter 7. Recapitulation and discussion 165
7.1. Background 165
7.2. Glycaemic variability in the prediction of cardiovascular complications 167
7.3. Lipid variability in the prediction of cardiovascular complication 169
7.4. Novelty and implications in the use of machine learning in cardiovascular risk stratification 171
7.5. Limitations 173
7.6. Future works 175
7.7. Conclusion 178
References 179
Appendices 196
Appendix 1. List of publications directly arising from the work described in this thesis 196

Appendix 2. Statement regarding published material submitted in this thesis and nature of contribution 198
Appendix 3. Correspondence from coauthor confirming contribution 200

Acknowledgements

Firstly, I would like to thank my family and my husband Justin for their support. Next, I would like to thank the team of clinicians and patients I have worked over the past years for their inspiration. Moreover, I would like to thank my supervisors- Professor Vassilios Vassiliou, Professor Johannes Reinhold for their kind supervision. Finally, I would like to thank Professor Gary Tse for his continuous generous guidance through my research career.

Abstract

Introduction: With the global trend of shifting towards personalised medicine, there is an increasing need for new parameters to individualise the disease-monitoring of type 2 diabetes mellitus (T2DM) beyond existing models such as the Reynolds Risk Score. Measures of glycaemic and lipid variability, defined as the extent of change in glycaemic/ lipid indices during follow-up, have attained academic interest as potential prognostic biomarkers in patients with T2DM and cardiovascular diseases. The present thesis aims to explore the use of glycaemic and lipid variability for predicting major adverse cardiovascular events amongst patients with T2DM.

Methods: A number of retrospective, population-based studies were included, which assessed the predictive values of glycaemic and lipid variability for various major adverse cardiovascular events. The study population included patients with T2DM attending the Hong Kong Hospital Authority between January 1st, 2009 till December 31st, 2009, with follow-up until December 31st, 2019. Demographic, clinical, biochemical and pharmacological data was extracted from a territory-wide, linked electronic database. Cox proportional hazards regression was applied with risk scores constructed from the hazard ratios. The models were further enhanced by machine-learning techniques.

Results: Up to 273 678 patients were analysed in the studies described herein. Glycaemic and lipid variability were found to be consistently predictive for major adverse cardiovascular events across the different studies ($p < 0.05$). HbA1c standard deviation ($p<0.0001$) and lipid indices (total cholesterol: $p=0.033$, high density lipoprotein: $p=0.082$) were found to be predictors of sudden cardiac death. Significant predictors of all-cause mortality were incorporated into a score-based predictive risk model that had a c-statistic of 0.73, which was improved to 0.86 (random survival forest) and 0.87 (deep survival learning models).

Conclusion: In conclusion, glycaemic and lipid variability can predict cardiovascular adverse events amongst patients with T2DM, allowing early intervention and management upon initial clinic visits in high-risk groups.

List of tables

Table 1. Baseline characteristics of predictions of diabetes complications and mortality using Hba1c variability: a 10-year observational cohort study 29
Table 2. Logistic regression of HbA1c value, variability and hypoglycaemia frequency on mortality and diabetes-related complications 30
Table 3 Cox regression of HbA1c value, variability and hypoglycaemia frequency on all-cause and cardiovascular mortality 32
Table 4. Discrete baseline characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning 48
Table 5. Continuous baseline characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning 49
Table 6. Adverse outcome characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning 52
Table 7. Biochemical predictor characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning 52
Table 8. Multivariate Cox regression of all-cause mortality in diabetes mellitus using machine learning 53
Table 9. Univariate Cox regression for adverse outcomes of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning 55
Table 10. Significant associations between HbA1c/ lipid variability with baseline neutrophil-lymphocyte ratio 67
Table 11. Model performance comparison analyses with five-fold cross validation 81
Table 12. Baseline characteristics for continuous variables for the development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong 94
Table 13. Baseline characteristics for discrete variables for the development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong 95
Table 14. Univariate predictors for all-cause mortality in diabetic patients in Hong Kong 97
Table 15. Multivariate predictors for all-cause mortality in diabetic patients in Hong Kong 100
Table 16. A score-based predictive risk model for all-cause mortality in type 2 diabetes mellitus 106
Table 17. Variable importance ranking generated by random survival forest model 109
Table 18. Survival prediction performance comparison between Cox, random survival forest and DeepSurv model with five-fold cross-validation approach 111
Table 19. Baseline clinical characteristics of risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 124
Table 20. Baseline biochemical characteristics of risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 125
Table 21. Univariate predictors for sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 126
Table 22. Univariate predictors for atrial fibrillation in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 127
Table 23. Univariate predictors for ventricular tachycardia in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 128
Table 24. Multivariate analysis showing predictors of sudden cardiac death (n=15316) in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 130
Table 25. Multivariate analysis showing predictors of atrial fibrillation (n=13267) in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study 131

Table 26. Baseline characteristics for categorical variables in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death 141
Table 27. Baseline characteristics for continuous variables in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death 143
Table 28. Univariate predictors for acute myocardial infarction in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death 145
Table 29. Multivariate predictors for acute myocardial infarction in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death 147
Table 30. Multivariate hazard ratios of acute myocardial infarction predictive score parameters 148
Table 31. Acute myocardial infarction prediction score 149
Table 32. Univariate predictors for sudden cardiac death in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death 150
Table 33. Multivariate predictors for sudden cardiac death in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death 152
Table 34. Multivariate hazard ratios of sudden cardiac death predictive score parameters 154
Table 35. Sudden cardiac death predictive score 155
Table 36. Variable importance ranking generated by conditional inference survival forest model 158
Table 37. Comparisons between conditional inference survival forest (CISF), multivariate Cox and random survival forest (RSF) model (five-fold cross validation) 160

List of figures

Figure 1. Plot of time-to-death of all-cause mortality against mean HbA1c using a generalized additive model....34
Figure 2. Kaplan-Meier plots of dichotomized mean HbA1c (A) and HbA1c variability (B) for all-cause mortality, dichotomized mean HbA1c (C) and HbA1c variability (D) for cardiovascular mortality....35
Figure 3. Workflow of regularized and weighted random survival forest model....48
Figure 4. Importance ranking and minimal depth of significant univariable variables to predict mortality and complications using regularized and weighted random survival forest model....69
Figure 5. Main tree-based decision rules to predict mortality and complications using regularized and weighted random survival forest model....77
Figure 6. Graphical representation of quartile hazard ratios from fasting blood glucose variability measures....99
Figure 7. Graphical representation of quartile hazard ratios from HbA1c variability measures....100
Figure 8. Graphical representation of hazard ratios for all-cause mortality....103
Figure 9: A: low-density lipoprotein-cholesterol (LDL-C); B: high-density lipoprotein-cholesterol (HDL-C), C: total cholesterol; D: Triglyceride. E: mean HbA1c. F: mean fasting blood glucose. The data points were generated by comparing the risks for mortality between patients in decile one and the other respective decile.....105
Figure 9. Receiver operating characteristic curve and area under the curve of the score-based predictive risk model for all-cause mortality....108
Figure 10. Selecting optimal tree number for random survival forest model....109
Figure 11. Kaplan-Meier survival curves of different age groups for sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study....123
Figure 12. Kaplan-Meier survival curves of different age groups for atrial fibrillation in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study....123
Figure 13. The association between mean HbA1c and acute myocardial infarction....146
Figure 14. The association between mean high-density lipoprotein cholesterol and acute myocardial infarction....147
Figure 15. The receiver operator characteristic curve for the acute myocardial infarction predictive score....150
Figure 16. The association between mean HbA1c and acute myocardial infarction....154
Figure 17. The association between mean total cholesterol and acute myocardial infarction....154
Figure 18. The receiver operator characteristic curve for sudden cardiac death predictive score....156
Figure 19. Optimal tree number selection for conditional inference survival model (five-fold cross validation) to predict for acute myocardial infarction....157
Figure 20. Optimal tree number selection for conditional inference survival forest model (five-fold cross validation) to predict for sudden cardiac death....157
Figure 21. Optimal tree number iteration and variable importance ranking generated by conditional inference survival forest model to predict acute myocardial infarction....159
Figure 22. Optimal tree number iteration and variable importance ranking generated by conditional inference survival forest model to predict sudden cardiac death....159

List of abbreviations

Abbreviations	Definition
ACCORDS	Action to Control Cardiovascular Risk in Diabetes
ACEI	angiotensinogen-converting-enzyme inhibitor
ADA	American Diabetes Association
ADVANCE	Action in Diabetes and Vascular Disease
AF	atrial fibrillation
ALP	alkaline phosphatase
ALT	alanine aminotransferase
AMI	acute myocardial infarction
ARB	angiotensin receptor blocker
AUC	area under the reactive operator characteristic
CCB	calcium channel blocker
CDARS	Clinical Data Analysis and Reporting System
CI	confidence interval
CISF	conditional inference survival forest
CKD	chronic kidney disease
CLD	chronic liver disease
COPD	chronic obstructive pulmonary disease
CV	coefficient of variation
DEVOTE	Comparing Cardiovascular Safety of Insulin Degludec vs Insulin Glargine in Patients with Type 2 Diabetes at High Risk of Cardiovascular Events
DKA	diabetic ketoacidosis
DPP4I	dipeptidyl peptidase-4 inhibitor
EASD	European Association for the Study of Diabetes
FBG	fasting blood glucose
GIP	glucose-dependent insulinotropic polypeptide
GLP-1A	glucagon-like peptide-1 receptor agonist
GRADE	The Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study
HAMDCT	Hospital Authority Master Disease Code Table
HDL-C	high-density lipoprotein cholesterol
hERG	human ether-a-go-go-related gen
HF	heart failure
HHS	hyperosmotic hyperglycaemia state
HR	hazard ratio
HVS	Hba1c variability score
ICD	International Classification of Diseases
ICH	intracranial haemorrhage
IHD	ischemic heart disease
IMPROVE-IT	Improved Reduction of Outcomes: Vytorin Efficacy International Trial
IQR	interquartile range
LDL-C	low-density lipoprotein cholesterol

MACE	major adverse cardiovascular events
MCH	mean corpuscular haemoglobin
MCHC	mean corpuscular haemoglobin concentration
MCV	mean corpuscular volume
NLR	neutrophil-lymphocyte ratio
OOB	out-of-bag
OR	odds ratio
ORIGIN	Outcome Reduction with Initial Glargine Intervention
OSLER-1/-2	Open-Label Study of Long-Term Evaluation against LDL Cholesterol -1/-2
PCSK9	proprotein convertase subtilisin/ kexin type 9
PPAR	peroxisome proliferator-activated receptor
PROMINENT	Pemafibrate to Reduce Cardiovascular Outcomes by Reducing Triglycerides in Patients with Diabetes
PVD	peripheral vascular disease
QTc	corrected QT interval
REDUCE-IT	Reduction of Cardiovascular Events with Icosapent Ethyl–Intervention Trial
RMS	root mean square
ROC	reactive operator characteristic
RSF	random survival forest
SCD	sudden cardiac death
SD	standard deviation
SGLT2I	sodium-glucose cotransporter-2 inhibitors
SNP	single nucleotide polymorphisms
STRENGTH	Long-Term Outcomes Study to Assess Statin Residual Risk with Epanova in High Cardiovascular Risk Patients with Hypertriglyceridaemia
T2DM	type 2 diabetes mellitus
TIA	transient ischemic attack
TNT	Treatment to New Targets
UKPDS	United Kingdom Prospective Diabetes Study
VADT	Veterans Affairs Diabetes Trial
VF	ventricular fibrillation
VT	ventricular tachycardia

Chapter 1. Introduction

1.1. Background

Major adverse cardiovascular events (MACE), including acute myocardial infarction, thromboembolic stroke, heart failure and peripheral vascular disease, are major contributors to morbidity and mortality amongst patients with type 2 diabetes mellitus (T2DM). The heightened risk of cardiovascular disease has been well-established for over five decades. A greater incidence of cardiovascular diseases amongst patients with diabetes mellitus, in comparison to their non-diabetic counterparts across all age groups, was first demonstrated by the Framingham Heart Study (1). Consequently, there has been significant academic interest in exploring ways to manage or reduce cardiovascular risk amongst patients with T2DM. One of the topics explored is the effects and extent of glycaemic or lipid control to lower the risks of MACE. In the past, the relationship between glycaemic or lipid control and cardiovascular risks was considered linear. However, recent evidence has shown that intensive glycaemic and lipid control may not be the best method to lower the risks for MACE in a safe manner (2). As a result, there has been a shift towards a more patient-centred, individualized approach in the long-term treatment of T2DM (3). Indeed, new, personalised disease-monitoring parameters were explored over the past decade (4, 5). Besides the absolute glycaemic and lipid concentrations, the temporal fluctuations of glycaemic and lipid control were identified as independent risk factors of increased cardiovascular disease burden amongst T2DM patients. Currently, glycaemic and lipid variability are yet to be introduced as routine disease-monitoring parameters amongst patients with T2DM. Hence, the present thesis aims to highlight the importance of temporal variability in glycaemic and lipid control, therefore change the status quo in the monitoring of cardiovascular risks in T2DM. In the present thesis, the use of glycaemic and lipid variability in the prediction of MACE in T2DM will be examined to improve the cardiovascular risk stratification amongst patients with T2DM.

1.2. The role of glycaemic control in cardiovascular risk control

Hyperglycaemia is a hallmark of T2DM, and an integral part of the pathogenesis of MACE amongst patients with T2DM. The diagnosis of T2DM is made by fulfilling any of the following criteria: 1) asymptomatic: a) fasting blood glucose (FBG) ≥ 7.0mmol/L; b) 2-hour blood glucose after 75g oral glucose tolerance test ≥ 11.1mmol/L; c) HbA1c ≥ 6.5%; 2) symptomatic of hyperglycaemia/ hyperglycaemia crisis: random blood glucose ≥ 11.1mmol/L (6). High HbA1c was reported to be a significant predictor for fatal and non-fatal cardiovascular diseases in a study of over 18 000 patients in the Swedish National Diabetes Register (7). Laboratory studies have shown that hyperglycaemia induces endothelial dysfunction and promotes atherogenesis, resulting in a greater atherosclerotic plaque burden and higher vulnerability to rupture (8). In addition, hyperglycaemia is also associated with higher levels of oxidative stress, thereby providing a pro-inflammatory and pro-thrombotic state. The systemic impact persists despite normalization of glucose concentrations, hence resulting in the significantly increased MACE risk amongst patients with T2DM (8, 9).

In the past, it was thought that the relationship between plasma glucose level and MACE was linear, but recent studies have reported U- or J-shaped relationships between measures of glycaemic control and MACE or its components (7). An example of such relationships between HbA1c or total cholesterol with MACE is illustrated in **Figure 1**. The United Kingdom Prospective Diabetes Study (UKPDS) was one of the first landmark studies that showed patients on more intensive glycaemic control to have a significantly lower risk for microvascular (but not macrovascular) adverse events (10). However, subsequent studies raised questions about the attainability of an intensive glycaemic target. In the Action to Control Cardiovascular Risk in Diabetes (ACCORDS) trial, the mortality rate was significantly higher in the group with intensive glycaemic control ($p = 0.04$) than their counterparts in the

standard glycaemic control group, with a greater incidence of significant weight gain and hypoglycaemia, resulting in early termination of the trial (11). It was found that persistent hypoglycaemia, albeit asymptomatic, may be pro-arrhythmic for patients with T2DM and high cardiovascular risks (12). In addition, a significant association between severe hypoglycaemia and all-cause mortality was established in the double-blind Comparing Cardiovascular Safety of Insulin Degludec vs Insulin Glargine in Patients with Type 2 Diabetes at High Risk of Cardiovascular Events (DEVOTE) study (13). As a result, an individualised approach has been adopted by recent guidelines. The 2022 Consensus from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD) states that whilst HbA1c ≥7% is a reasonable glycaemic target for most non-pregnant adults, a lower HbA1c target could be sought for if it can be attained safely without adverse health effects, and a higher target is appropriate for patients with limited life expectancy, poor premorbid state or advanced complications (14).

The introduction of novel classes of anti-diabetic agents has provided additional options for achieving glycaemic control safely with protective cardiovascular effects. Sodium-glucose cotransporter-2 inhibitors (SGLT2I) are the newest antidiabetic agent that lowers blood glucose by promoting urinary glucose excretion. Significant cardio-renal protective effects were demonstrated across different clinical trials (15, 16), with the greatest benefit in the reduction of hospitalization for heart failure (HF) and renal outcomes (17). Interestingly, the cardiovascular-protective effects of SGLT2I may derive not only urinary glucose excretion, but also mechanisms that are independent of their glucose-lowering effects. Although not completely understood, the combination of preload and afterload reduction, attenuation of cardiac fibrosis, and improvements in myocardial metabolism may directly contribute to the reduction of adverse events (18). Although older studies raised concerns over the complications of euglycaemic diabetic ketoacidosis and urinary tract infection amongst SGLT2I users (19),

the absolute risk is low and can be mitigated by patient education (20) and temporary withholding of the medications in the presence of conditions that can predispose to ketoacidosis. Gradually, SGLT2Is are now recommended for populations beyond T2DM (21), with real-world evidence reporting lower risks of not only adverse cardiovascular events (22, 23), but also other adverse outcomes (24, 25) as well as randomised studies and meta-analyses showing these beneficial effects (26, 27).

Glucagon-like peptide-1 receptor agonist (GLP-1A) is another new class of antidiabetic agents with potent glucose-lowering and weight-reducing effects. In addition to augmenting the secretion of insulin and suppression of glucagon, GLP-1A slows gastric emptying, resulting in the curbing of postprandial hyperglycaemia and improving appetite control, thus improving the attainment of glycaemic targets (28, 29). In addition, GLP-1A has been reported to reduce MACE of T2DM patients with established cardiovascular diseases, or high cardiovascular risk (30). Liraglutide has been reported to be more effective in achieving the target HbA1c in comparison to a sulphonylurea or dipeptidyl peptidase-4 inhibitor (DPP4I) with a lower risk in the composite outcome of MACE, revascularization, and heart failure/ unstable angina requiring hospitalization by The Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study (GRADE) multicentre open-label randomised controlled trial (31). Tirzepatide, a combination of a glucose-dependent insulinotropic polypeptide (GIP) and GLP-1A approved by the United States Food and Drug Administration in 2022, demonstrated superior glucose-lowering effects than other long-acting GLP-1A (14). Current preliminary data shows that tirzepatide has comparable MACE-lowering effects to its long-acting GLP-1A counterparts, with ongoing trials on its long term cardiovascular profile (32).

1.3. The role of lipid control in cardiovascular risk control

Low-density lipoprotein cholesterol (LDL-C) drives atherogenesis via a multitude of mechanisms including the induction of an endothelial inflammatory response and promotion of plaque rupture. Besides reducing the formation of atherosclerotic plaques, studies have shown that the lowering of LDL-C stabilizes existing plaques by changing the plaque contents (33). By increasing the thickness of the fibrous cap, the atheroma has a lower risk of rupture and subsequent thrombosis, therefore ultimately reducing the risk of MACE (34). Similarly, elevated triglyceride also marks an increased risk for MACE in patients with T2DM since it represents both the concentration of atherogenic remnant cholesterols in the circulation and the tissue resistance against insulin (35). The use of lipid-lowering therapy can lead to the regression of carotid artery stenosis, suggesting that intensive lipid control may be able to halt or even reverse the progression of atherosclerotic cardiovascular disease (34). As a result, the attainment of lipid control has become an integral part of the treatment goal for T2DM.

Advancements in strategies of lipid control played an important role in the improvement of cardiovascular risk reduction amongst T2DM patients. In a meta-analysis that included 14 randomised controlled trials on the use of lipid-lowering agents amongst patients with T2DM, it was shown that a 1 mmol/L reduction in LDL-C reduces the risk of major vascular events by 21% (36). Recent evidence shows that aggressive LDL-C control can attain a further reduction in cardiovascular risk. As reported by the Treatment to New Targets (TNT) study, lowering LDL-C to < 1.99 mmol/L, below the recommended LDL-C threshold of 2.59 mmol/L at the time, resulted in a 25% reduction in the risk of MACE (37).

In the past, statins were the major lipid-lowering agents with benefits derived from both their lipid-lowering and anti-inflammatory effects as well as presumed “pleiomorphic” effects. Whilst initially there was some evidence to suggest a modest increase in the risk of incident T2DM amongst statin users, further work has shown this to relate predominantly to existing

metabolic risk factors and ageing. A meta-analysis that included over 90 000 participants from 13 trials reported a 9% increase in risk for T2DM over a mean course of four years (38).

With the development of non-statin lipid-lowering strategies, more aggressive LDL-C targets became achievable. In the Improved Reduction of Outcomes: Vytorin Efficacy International Trial (IMPROVE-IT) study, ezetimibe was able to further lower LDL-C by 24% in patients with relatively lower baseline LDL-C levels, with greater cardiovascular protective effects noted amongst users with T2DM (39). Bempedoic acid, a promising novel oral adenosine triphosphate-lyase inhibitor that inhibits upstream to 3-hydroxy-3-methylglutaryl coenzyme A reductase, was shown to reduce the risk of atherosclerotic cardiovascular disease by down-regulating pro-inflammatory pathways in a pooled analysis of four recent phase 3, double-blind, randomised controlled trials (40). In another double-blinded, randomised, placebo-controlled trial, bempedoic acid is shown to reduce the risk of MACE in statin-intolerant patients (41). Since bempedoic acid is activated in the liver, instead of in peripheral muscles, it may have contributed to its lower risks for musculoskeletal adverse effects. Inclisiran, which inhibits the synthesis of proprotein convertase subtilisin-kexin type 9 (PCSK9) in the liver, is able to reduce LDL-C level by approximately 50% in patients on maximally tolerated statin doses in two phase 3 trials (42).

Monoclonal antibodies to PCSK9 are the newest class of non-statin lipid-lowering agents, and are used as an add-on therapy to patients with high cardiovascular risks on maximally tolerated statin therapy, in patients with familial hypercholesterolemia, or as an alternative to statins in statin-intolerant patients (43). In the Open-Label Study of Long-Term Evaluation against LDL Cholesterol -1/-2 (OSLER-1/-2) trials, it was reported that the risk for MACE was reduced dramatically by more than half at one-year follow-up (44). Evidence for the cardiovascular protective effects of PCSK9 inhibitors is preliminary, but promising, and may be related to its effects on the immune system (45, 46).

By contrast, evidence on the cardiovascular-protective effects of triglyceride-lowering therapies is more limited. Fibrates, which lowers triglyceride levels by peroxisome proliferator-activated receptor (PPAR) modulation, can effectively reduce triglyceride levels up to 70%, albeit with significant individual variations (35). However, clinical trials noted neutral effects of fibrate in the reduction of atherosclerotic cardiovascular disease and do not provide further MACE risk reduction when used in combination with a statin (47). The recent Pemafibrate to Reduce Cardiovascular Outcomes by Reducing Triglycerides in Patients with Diabetes (PROMINENT) multinational, double-blinded, randomised controlled trial showed that pemafibrate, a potent selective PPAR-alpha modulator, does not lower the incidence of MACE despite its efficacy in the reduction of triglyceride, very-low density lipoprotein-cholesterol, remnant cholesterol, and apolipoprotein C-III levels (48). Therefore, the use of fibrates remains to be reserved for patients with isolated or persistent hypertriglyceridaemia despite optimally controlled LDL-C. Additionally, evidence for the cardiovascular-protective effects of isosapent ethyl, a purified omega-3 fatty acid, remained controversial. Although it was previously shown to be able to reduce the incidence of atherosclerotic cardiovascular diseases in high doses by the Reduction of Cardiovascular Events with Icosapent Ethyl–Intervention Trial (REDUCE-IT) trial (49), the subsequent Long-Term Outcomes Study to Assess Statin Residual Risk with Epanova in High Cardiovascular Risk Patients with Hypertriglyceridaemia (STRENGTH) trial, which uses corn oil instead of mineral oil as placebo, failed to show the same cardioprotective effects (50). Therefore, contrary to LDL-C management, the control of hypertriglyceridaemia is not a prioritised treatment target.

1.4. Role and contributors to temporal variability

With a shift towards a personalised approach in the management of T2DM, there has been a call for novel, individualised disease-monitoring parameters in the evaluation of

cardiovascular risks among patients. The predictive value of time-varying parameters in glycaemic and lipid control has become a topic of interest over the past decade.

Independent of HbA1c and glucose levels, increased glycaemic variability has been reported to increase the risk of microvascular, and macrovascular complications and MACE (51). A meta-analysis of 13 studies demonstrated that long-term glycaemic variability, represented by HbA1c variability, is associated with a higher risk of cardiovascular disease, macrovascular events, renal disease and all-cause mortality (8). In a *post-hoc* analysis of the visit-to-visit HbA1c variability and fasting glucose of the Action in Diabetes and Vascular Disease (ADVANCE) trial, raised fasting glucose and HbA1c variability are both associated with an increased risk for macrovascular adverse events (52). Similarly, a *post-hoc* analysis of the Veteran Affairs Diabetes Trial (VADT) demonstrated a significant positive association between fasting glucose variability and cardiovascular disease after adjusting for risk factors, including mean fasting glucose (53). Machine-learning techniques were applied recently to improve the accuracy of predictive models by accounting for the interactions between cardiovascular risk factors (54).

Besides long-term glycaemic variability, the predictive value of short-term day-to-day variability has also been explored. The DEVOTE trial showed that the standard deviation of the self-monitored blood glucose over three days was associated with MACE and all-cause mortality (51). Both day-to-day FBG and HbA1c variability were associated with severe hypoglycaemia (55), possibly suggesting hypoglycaemia as a mediator for MACE.

By contrast, the evidence supporting the relationship between lipid variability and increased cardiovascular risk is less consistent. A recent meta-analysis including 11 studies from 7 cohorts of the general population shows that those with top quartile total cholesterol, high-density lipoprotein cholesterol (HDL-C), and LDL-C have a higher risk for cardiovascular disease and all-cause mortality. (56) However, high triglyceride variability is not associated

with an increased cardiovascular risk. In terms of patients with T2DM, raised LDL-C and triglyceride variability were reported to be associated with an elevated cardiovascular risk, particularly amongst young patients between ages 45-54 years old. (57)

1.5. Conclusion

To conclude, glycaemic and lipid control is an integral part of disease management in T2DM, particularly in terms of the control of risk for MACE. With the call for an individualised approach in the management of T2DM, novel disease-monitoring parameters were explored. The temporal variability of glycaemic and lipid parameters was noted to be of significant predictive value in the risk stratification of cardiovascular diseases, particularly amongst patients with T2DM. Whilst more evidence was needed to elucidate the pathogenic mechanism underlying high glycaemic and lipid variability on increased cardiovascular risk, temporal variability of glucose and lipid control were being incorporated into cardiovascular risk stratification models to improve the accuracy of predictions.

Chapter 2. Predictions of diabetes complications and mortality using Hba1c variability: a 10-year observational cohort study

2.1. Introduction

T2DM is an increasingly prevalent metabolic disease with a significant global disease burden. Currently, it affects more than 400 million individuals across the globe, with the number of affected patients projected to increase by more than 50% by 2045, and an age-dependent increase in prevalence (1). Although normalization of blood glucose remains to be the treatment goal for diabetic patients, the extent of glycaemic control remains controversial. Previous large-scale clinical trials, such as the UKPDS and the ADVANCE trial, have reported a significant reduction in mortality and cardiovascular complications by tighter glycaemic control excluded patients with major comorbidities (2, 3, 5). Besides, the 2008 ACCORD trial ended prematurely due to significantly higher mortality reported in the intensive glycaemic control group (58). There is increasing evidence for increased mortality risk for patients in both extremes of HbA1c, which drives for less stringent glycaemic control for the elderly population (4, 6, 7). However, the lower limits of glycaemic control have yet to be clearly outlined in the current guidelines (59). Additionally, research has shifted to exploring other parameters that can facilitate more individualized disease-monitoring.

Emerging evidence suggests that HbA1c variability, in addition to HbA1c itself, can be used as a predictor for complications and mortality. Although the underlying mechanism remains unclear, increased HbA1c variability has been associated with diabetic complications in various organ systems, in addition to all-cause and cardiovascular mortality (8, 60, 61, 62). Different theories have been proposed to explain the association, including that a wide variance in HbA1c may reflect higher complexity in the disease course, suboptimal management, and poorer baseline vascular conditions(63, 64). Other investigators have proposed the involvement of intermittent hypoglycaemia, where the resulting increased oxidative stress and sympathoadrenal activation induce additional stress on end organs under chronic inflammation

(12, 65, 66). However, there is a lack of evidence from large-scale clinical studies to support the hypothesis.

The present study aims to examine the predictive power of both HbA1c value and variability towards the prognosis of diabetic patients. Furthermore, to test the hypothesis that intermittent hypoglycaemia underlies the predictive value of HbA1c variability towards the prognosis of diabetic patients, the inter-relationship between hypoglycaemia, HbA1c variability, and mortality will be evaluated.

2.2. Methods

2.2.1. Study population

The single-centre retrospective observational study was approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee. The present study consists of patients with T2DM prescribed insulin at outpatient clinics of the Prince of Wales Hospital and Shatin Hospital from January 1st, 2009 to December 31st, 2009. The patients were identified from the Clinical Data Analysis and Reporting System (CDARS), a territory-wide healthcare database.

2.2.2. Dataset description

CDARS is a territory-wide database that centralizes patient information from individual local hospitals, general and specialist outpatient clinic under the Hong Kong Hospital Authority to establish comprehensive medical data, including clinical characteristics, disease diagnosis, laboratory results, and drug treatment details. Since the healthcare database in Hong Kong were only shifted to an electronic system after the year 1999, prior records were not available in the system. Laboratory results and drug prescriptions were automatically uploaded to CDARS.

Disease diagnosis were documented in terms of International Classification of Diseases Nineth Edition (ICD-9) coding in discharge summaries of inpatient admissions or consultation notes in clinic visits. CDARS is also linked to the death registry for the extraction of mortality outcomes, with causes of death documented in ICD-10 codings. Lifestyle factors, such as smoking status or body mass index, were not available. Free text was not captured by the system. The system has been previously used by both our team and other teams in Hong Kong (67, 68). Patients with three or more HbA1c measurements were included in the analysis for HbA1c variability.

2.2.3. Patient data

Clinical and biochemical data were extracted for the present study. Data on the primary outcomes, all-cause and cardiovascular mortality, between January 1st, 2009 to May 1st, 2019 was obtained. Data on secondary outcomes between January 1st, 2009 to December 31st, 2013 were extracted, including 1) neurological, ophthalmological and renal diabetic complications, 2) microalbuminuria and macroalbuminuria, 3) peripheral vascular disease (PVD), 4) stroke and transient ischemic attack (TIA), 5) atrial fibrillation (AF), 6) sudden cardiac death (SCD), 7) diabetic ketoacidosis or hyperosmotic hyperglycaemia state (DKA/ HHS) or coma.

Microalbuminuria was defined as fulfilling any of the following: 1) urine albumin/ creatinine ratio between 3mg/mmol to 30mg/mmol, 2) 24-hour total urine albumin between 30mg/ day to 300mg/ day, 3) spot urine albumin between 30mg/L to 300mg/L. Macroalbuminuria is defined as 1) urine albumin/ creatinine ratio > 30mg/mmol, 2) 24-hour total urine albumin >300mg/ day, 3) spot urine albumin >300mg/L. Proteinuria was defined as either 24-hours total urine protein >3.5g/day, or albumin/ creatinine ratio >30mg/mmol. SCD was defined as the occurrence of ventricular tachyarrhythmia or non-specific cardiac arrest. Patients with established events before recruitment for a given outcome were excluded. For

other outcomes, if there were no events, then these patients were included. Cardiovascular mortality was recorded using ICD-10 coding, whilst the remaining outcomes were documented in CDARS under ICD-9 codes.

Furthermore, baseline clinical details include 1) age, 2) sex, 3) specific comorbidities (chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD), chronic liver disease (CLD), HF, ischemic heart disease (IHD), hypertension, AMI, stroke). To capture the episodic occurrence of diseases, such as AMI, or the initiation of chronic conditions, such as hypertension, data on patient diagnosis from January 1st, 1999 to December 31st, 2008 was extracted. The patient's age is defined as age on January 1st, 2009. Additionally, the dosing regimen of antidiabetic and cardiovascular medications prescribed were extracted. The mean daily dose, which is the product between the daily dosing frequency and dosage, is reported for each drug class. The classes of anti-diabetic agents include: 1) insulin, 2) sulphonylurea, 3) biguanide, 4) alpha-glucosidase inhibitor, 5) thiazolidinedione, 6) meglitinide, 7) dipeptidyl peptidase-4 inhibitor (DPP4I), 8) GLP-1A. The cardiovascular medications include 1) angiotensinogen-converting-enzyme inhibitor/ angiotensin receptor blocker (ACEI/ARB), 2) beta-adrenergic receptor blocker, 3) calcium channel blocker (CCB), 4) diuretics.

Baseline biochemical data, defined as urinalysis or blood test results measured from January 1st, 2008 to December 31st, 2008, were extracted. Urinalysis results include 1) albumin/ creatinine ratio, 2) creatinine clearance, 3) spot protein, albumin, and glucose, 4) 24-hour total protein, and albumin. Data from renal function test, liver function test, and other baseline blood tests include 1) serum creatinine, 2) serum sodium, potassium, urate and urea, 3) serum albumin, 4) serum total protein, 5) serum total bilirubin, 6) serum alanine aminotransferase (ALT), 7) alkaline phosphatase (ALP), 8) FBG and random blood glucose, 9) total, HDL-C, direct and calculated LDL-C, 10) triglyceride. The following results were obtained from complete blood count: 1) haemoglobin, 2) mean corpuscular haemoglobin (MCH), 3) mean corpuscular

haemoglobin concentration (MCHC), 4) mean corpuscular volume (MCV), 5) haematocrit, 6) basophil count, 7) eosinophil count, 8) lymphocyte count, 9) monocyte count, 10) neutrophil count, 11) platelet count, 12) red cell count, 13) white cell count. HbA1c from January 1st, 2004 to December 31st, 2008 was extracted to establish the baseline HbA1c and HbA1c variability. Random and FBG were also extracted to obtain the episodes of hypoglycaemia, defined by blood glucose < 3.9 mg/mmol. Results from the latest test that took place in 2008 were used as the baseline if multiple tests were performed during the year.

2.2.4. Statistical analysis

Continuous variables were presented as mean ± standard deviation. HbA1c temporal variability was examined through the following approaches: 1) mean, 2) standard deviation (SD), 3) root mean square (RMS), 4) coefficient of variation (CV). RMS is calculated by first squaring all HbA1c values, then square root the mean of the squares. The CVwas obtained by dividing the HbA1c SD by the mean HbA1c, expressed as a percentage. Hba1c variability score (HVS) was defined as the number of HbA1c measurements > 0.5% of the previous reading divided by the total number of HbA1c measurements, expressed as a percentage.

Logistic regression was used to identify significant predictors of the different outcomes. Cox regression was applied to evaluate the predictive value of HbA1c variability for time-to-death in all-cause mortality and cardiovascular mortality. Time-to-death was defined by the number of days from January 1st, 2009 to the date of death of the patient, or until May 1st, 2019. 95% confidence interval (CI) was presented in the analyses, with odds ratio (OR) and hazard ratio (HR) for logistic and Cox regression respectively.

Several methods were adopted to further outline the relationship between blood glucose value, HbA1c variability, and mortality. The relationship between mean HbA1c and time-to-death for both all-cause and cardiovascular mortality was modelled using the generalized

additive model, with 95% CI displayed. Cut-off values of HbA1c mean and SD for the prediction of all-cause and cardiovascular mortality were derived by the maximization of sensitivity and specificity, using the Liu method, and the area under the reactive operator characteristic (ROC) curve (AUC) was calculated. SD was used to represent HbA1c variability since it is the only parameter unrelated to the value of HbA1c among the four parameters for variability.

The cohort was then dichotomized into "high" and "low" HbA1c value and variability based on the HbA1c mean and SD respectively. Kaplan-Meier curves were plotted for the dichotomized cohorts against the time-to-death for all-cause mortality, with the significance of intergroup differences assessed by the log-rank test. To elucidate the underlying connections between intermittent hypoglycaemia, HbA1c variability, and mortality, Poisson regression was used to assess the relationship between hypoglycaemia frequency and dichotomized HbA1c variability, whilst both logistic and Cox regression were used to test the predictive ability of hypoglycaemia frequency towards both all-cause and cardiovascular mortality. Statistical significance is defined as P-value <0.05. All statistical analyses were performed using R Studio.

2.3. Results

2.3.1. Clinical and biochemical characteristics

The present cohort consists of 3424 patients (median age= 63, interquartile range of age= 20 years, male= 50.2%). The baseline biochemical parameters of the cohort are presented in **Table 1**. Within the present cohort, the most common comorbidity is hypertension (24.6%), followed by IHD (15%), stroke (11%), HF (10%), CKD (7%), CLD (5%), AMI (4%) and COPD (3%). In terms of drug prescriptions, the mean daily insulin dose was 20.8 ± 13.0 units. Other classes of antidiabetic agents prescribed include 1) biguanide (n=1300, mean daily dose=

1546 ± 742mg); 2) sulphonylurea (n=1300, mean daily dose= 131 ± 122mg); 3) thiazolidinedione (n= 268, mean daily dose= 7.17 ± 8.93mg); 4) alpha-glucosidase inhibitor (n= 89, mean daily dose= 185 ± 80.0mg); 5) GLP-1A (n=8, mean daily dose= 11.8 ± 4.05mg); 6) DPP4I (n=7, mean daily dose= 97.8 ± 14.9mg). The following cardiovascular medications were prescribed: 1) ACEI/ ARB (n=2328, mean daily dose= 20.6 ± 38.9mg); 2) CCB (n=1579, mean daily dose= 58.2 ± 49.0mg); 3) beta-adrenergic receptor blockers (n= 1425, mean daily dose= 79.3 ± 64.3mg); 4) diuretics (n=955, mean daily dose= 66.3 ± 82.2mg).

Table 1. Baseline characteristics of predictions of diabetes complications and mortality using Hba1c variability: a 10-year observational cohort study

	Mean/Median/n	Standard Deviation/ Interquartile Rate/%
Demographic		
Age	60	20
Male	1718	50.2
Urinalysis		
Albumin/Creatinine Ratio (mg/mmol)	52.2	174
Creatinine Clearance (ml/min)	45.4	39.5
Spot Protein (g/d)	1.67	2.69
Spot Albumin (mg/L)	256.3	744
Spot Glucose (mmol/L)	11.4	6.10
24-hours Total Protein (g/d)	1.30	1.93
24-hours Total Albumin (mg/d)	324	715
Baseline Blood Test		
Fasting Glucose (mmol/L)	8.65	3.59
Random Glucose (mmol/L)	11.5	6.14
HbA1c (%)	8.05	1.66
Total Cholesterol (mmol/L)	4.60	1.11
High Density Lipoprotein (HDL) Cholesterol (mmol/L)	1.36	0.421
Low Density Lipoprotein (LDL) Cholesterol (Calculated) (mmol/L)	2.67	1.15
LDL Cholesterol (Direct) (mmol/L)	2.45	0.860
Triglyceride (mmol/L)	1.84	1.96
Renal Function Test		
Creatinine (umol/L)	147	167
Sodium (mmol/L)	139	2.95
Potassium (mmol/L)	4.28	0.461
Urate (umol/L)	397	124
Urea (mmol/L)	8.88	5.91
Liver Function Test		

Albumin (g/L)	40.8	4.60
Alanine Aminotransferase (ALT) (U/L)	26.0	21.0
Alkaline Phosphatase (ALP) (U/L)	80.8	41.0
Total Bilirubin (umol/L)	11.6	7.35
Total Protein (g/L)	78.0	6.24

2.3.2. HbA1c variability and baseline

Within the study cohort, 3137 patients had at least three HbA1c measurements, and the average number of HbA1c measurements per patient was 11.9 ± 4.8. The average frequency of hypoglycaemia was 0.6 ± 1.3 episodes. Throughout the study period (January 1st, 2009 to December 31st, 2019), there were 1491 cases of all-cause mortality, of which 308 were attributed to cardiovascular causes. Overall, the mean baseline HbA1c is 8.1 ± 1.8% (interquartile range= 2.0%), with HbA1c variability represented by 1) patient-specific mean (8.0 ± 1.2%); 2) SD (1.1 ± 0.71%); 3) RMS (8.1 ± 1. 2%); 4) CV (13.6 ± 7.6%). The logistic and Cox regression analysis for outcome prediction is presented in **Table 2** and **Table 3** respectively.

Table 2. Logistic regression of HbA1c value, variability and hypoglycaemia frequency on mortality and diabetes-related complications

	Odds Ratio (OR)	95% Confidence Interval (CI)	P-Value
Neurological Complications (n=3236)			
Baseline	1.07	[1.00, 1.14]	**0.033**
Mean	1.16	[1.05, 1.28]	**0.003**
Standard Deviation	1.23	[1.07, 1.42]	**0.004**
Root Mean Square	1.15	[1.05, 1.27]	**0.002**
Coefficient of Variation	1.02	[1.01, 1.03]	**0.006**
Ophthalmological Complications (n=3023)			
Baseline	1.10	[1.05, 1.15]	**<0.001**
Mean	1.32	[1.23, 1.42]	**<0.001**
Standard Deviation	0.990	[0.869, 1.12]	0.880
Root Mean Square	1.28	[1.19, 1.37]	**<0.001**
Coefficient of Variation	0.989	[0.977, 1.00]	0.089
Renal Complications (n=2973)			

Baseline	1.00	[0.956, 1.05]	0.843
Mean	1.11	[1.03, 1.19]	**0.006**
Standard Deviation	1.14	[1.02, 1.28]	**0.025**
Root Mean Square	1.10	[1.03, 1.18]	**0.007**
Coefficient of Variation	1.01	[1.00, 1.02]	0.051
Microalbuminuria and Macroalbuminuria (n=1912)			
Baseline	1.04	[0.976, 1.10]	0.236
Mean	1.18	[1.08, 1.29]	**<0.001**
Standard Deviation	1.01	[0.859, 1.17]	0.918
Root Mean Square	1.16	[1.07, 1.27]	**<0.001**
Coefficient of Variation	0.992	[0.976, 1.01]	0.286
Proteinuria (n=2572)			
Baseline	1.05	[0.982, 1.11]	0.148
Mean	1.25	[1.14, 1.37]	**<0.001**
Standard Deviation	1.15	[0.986, 1.32]	0.064
Root Mean Square	1.23	[1.13, 1.34]	**<0.001**
Coefficient of Variation	1.01	[0.993, 1.02]	0.257
Peripheral Vascular Disease (n=3375)			
Baseline	1.11	[0.999, 1.23]	**0.041**
Mean	1.05	[0.876, 1.25]	0.576
Standard Deviation	1.10	[0.814, 1.39]	0.489
Root Mean Square	1.05	[0.877, 1.24]	0.605
Coefficient of Variation	1.01	[0.984, 1.04]	0.387
Stroke (n=3168)			
Baseline	0.997	[0.932, 1.06]	0.932
Mean	1.09	[0.989, 1.20]	0.080
Standard Deviation	1.06	[0.904, 1.23]	0.429
Root Mean Square	1.08	[0.987, 1.19]	0.089
Coefficient of Variation	1.01	[0.990, 1.02]	0.473
Transient Ischemic Attack (n=3361)			
Baseline	1.06	[0.903, 1.22]	0.440
Mean	1.02	[0.792, 1.29]	0.888
Standard Deviation	0.795	[0.463, 1.22]	0.359
Root Mean Square	1.00	[0.785, 1.25]	0.985
Coefficient of Variation	0.983	[0.937, 1.02]	0.425
Atrial Fibrillation (n=3246)			
Baseline	1.01	[0.941, 1.08]	0.739
Mean	0.927	[0.828, 1.03]	0.183
Standard Deviation	0.858	[0.693, 1.04]	0.143
Root Mean Square	0.923	[0.827, 1.03]	0.147
Coefficient of Variation	0.986	[0.967, 1.00]	0.134
Sudden Cardiac Death (n=3408)			
Baseline	1.05	[0.961, 1.14]	0.274
Mean	0.939	[0.826, 1.06]	0.323
Standard Deviation	1.06	[0.865, 1.27]	0.544
Root Mean Square	0.943	[0.834, 1.06]	0.342
Coefficient of Variation	1.01	[0.991, 1.03]	0.297

Diabetic Ketoacidosis/ Hyperosmotic Hyperglycaemia State (DKA/ HHS)(n=3249)			
Baseline	1.16	[1.04, 1.28]	**0.006**
Mean	1.32	[1.14, 1.52]	**<0.001**
Standard Deviation	1.27	[1.02, 1.53]	**0.018**
Root Mean Square	1.29	[1.13, 1.47]	**<0.001**
Coefficient of Variation	1.02	[0.995, 1.04]	0.112
All-Cause Mortality (n=3424)			
Baseline	0.887	[0.848, 0.927]	**<0.001**
Mean	0.898	[0.845, 0.953]	**<0.001**
Standard Deviation	1.34	[1.21, 1.49]	**<0.001**
Root Mean Square	0.918	[0.866, 0.973]	**0.004**
Coefficient of Variation	1.03	[1.02, 1.04]	**<0.001**
Hypoglycaemia Frequency	1.11	[1.05, 1.17]	**<0.001**
Cardiovascular Mortality (n=3424)			
Baseline	0.889	[0.817, 0.964]	**0.005**
Mean	0.985	[0.886, 1.09]	0.779
Standard Deviation	1.42	[1.23, 1.66]	**<0.001**
Root Mean Square	1.02	[0.918, 1.12]	0.757
Coefficient of Variation	1.03	[1.02, 1.05]	**<0.001**
Hypoglycaemia Frequency	1.08	[0.990, 1.16]	0.070

The baseline, mean, SD/RMS/CV of HbA1c and hypoglycaemic frequency are used to predict for mortality and other complications in individual univariate logistic regression models.

Table 3 Cox regression of HbA1c value, variability and hypoglycaemia frequency on all-cause and cardiovascular mortality

	Hazard Ratio (HR)	95% Confidence Interval (CI)	P-Value
All-Cause Mortality			
Baseline	0.917	[0.886, 0.949]	**<0.001**
Mean	0.908	[0.868, 0.951]	**<0.001**
Standard Deviation	1.21	[1.14, 1.28]	**<0.001**
Root Mean Square	0.925	[0.885, 0.967]	**<0.001**
Coefficient of Variation	1.02	[1.02, 1.03]	**<0.001**
Hypoglycaemia Frequency	1.08	[1.04, 1.11]	**<0.001**
Cardiovascular Mortality			
Baseline	0.889	[0.817, 0.964]	**0.005**
Mean	0.984	[0.893, 1.09]	0.748
Standard Deviation	1.34	[1.21, 1.48]	**<0.001**
Root Mean Square	1.01	[0.924, 1.11]	0.773
Coefficient of Variation	1.03	[1.02, 1.04]	**<0.001**
Hypoglycaemia Frequency	1.07	[1.00, 1.15]	**0.044**

The baseline, mean, SD/RMS/CV of HbA1c and hypoglycaemic frequency are used to predict for mortality and other complications in individual univariate Cox regression models.

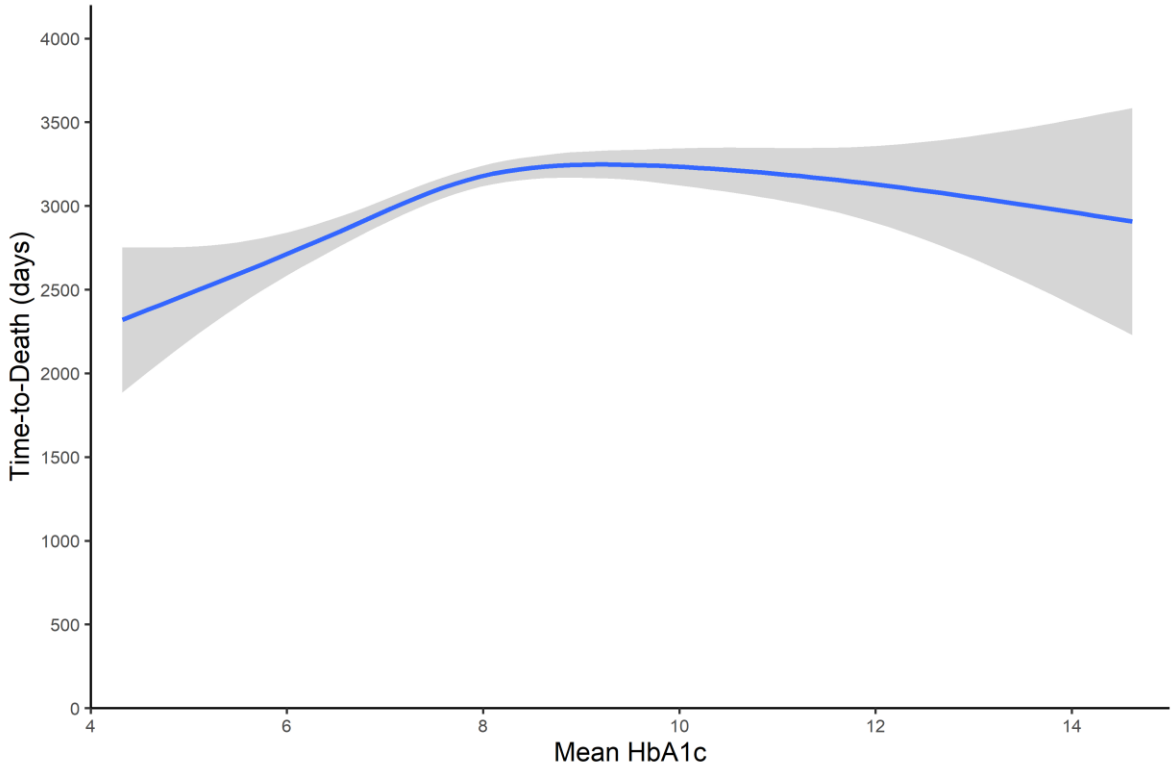
For the prediction of the primary outcomes- all-cause and cardiovascular mortality, both baseline HbA1c and HbA1c variability were found to be significant predictors. Interestingly, under logistic regression, baseline (OR= 0.89, 95% CI=[0.85, 0.93], P <0.001), mean (OR= 0.90, 95% CI=[0.85, 0.95], P <0.001), and RMS (OR= 0.92, 95% CI=[0.87, 0.97], P <0.001) of HbA1c were found to be negative predictors of all-cause mortality, whilst SD (OR= 1.34, 95% CI=[1.21, 1.49], P <0.001), CV (OR= 1.03, 95% CI=[1.02, 1.04], P <0.001) and HVS (OR= 1.007, 95% CI=[1.004, 1.009], P<0.001) were found to be significant positive predictors. Similar findings were found under Cox regression, with baseline (HR=0.92, 95% CI= [0.89, 0.95], P < 0.001), mean (HR=0.91, 95% CI= [0.87, 0.95], P < 0.001) and RMS (HR=0.93, 95% CI= [0.89, 0.97], P < 0.001) as negative predictors, SD (HR=1.21, 95% CI= [1.14, 1.28], P < 0.001), CV (HR=1.02, 95% CI= [1.02, 1.03], P < 0.001) and HVS (HR=1.01, 95% CI=[1.01, 1.01], P<0.001) as positive predictors for time-to-death.

Similar patterns were observed in cardiovascular mortality. Whilst baseline HbA1c is a negative predictor for both cardiovascular-specific mortality (OR= 0.89, 95% CI= [0.82, 0.96], P= 0.005) and time-to-death (HR=0.90, 95% CI= [0.84, 0.98], P= 0.009), SD (mortality: OR= 1.42, 95% CI= [1.23, 1.66], P < 0.001; time-to-death: HR= 1.34, 95% CI= [1.21, 1.48], P < 0.001), CV (mortality: OR= 1.03, 95% CI= [1.02, 1.05], P < 0.001; time-to-death: HR= 1.03, 95% CI= [1.02, 1.04], P < 0.001) and HVS (mortality: HR= 1.02, 95% CI= [1.01, 1.03], P<0.001; time-to-death: HR=1.02 95% CI: [1.01, 1.03], P<0.001) are positive predictors.

The individual effects of HbA1c and blood glucose values, in addition to HbA1c variability, on all-cause and cardiovascular mortality were further examined. The shorter time-to-death at the extremes of mean HbA1c in the generalized additive model for all-cause mortality is presented in **Figure 1**. The cut-off values for dichotomization of HbA1c value and variability were 7.3% (AUC= 0.540) and 0.86 (AUC= 0.574) in all-cause mortality prediction, and 6.8% (AUC= 0.493) and 0.88 (AUC= 0.590) for cardiovascular mortality prediction. After

dichotomization, the low mean HbA1c subgroup showed a significantly shorter time-till-death for all-cause mortality, (P < 0.001), but not cardiovascular mortality (P= 0.920). By contrast, the time-till-death was significantly shorter for the high HbA1c variability subgroup for both all-cause (P < 0.001), and cardiovascular mortality (P < 0.001).

Figure 1. Plot of time-to-death of all-cause mortality against mean HbA1c using a generalized additive model



The Kaplan-Meier plots for the dichotomized HbA1c value and variability in the prediction of all-cause and cardiovascular mortality are presented in **Figure 2**. A significant association was found between dichotomized HbA1c variability and hypoglycaemia frequency (P < 0.0001). Hypoglycaemia frequency was found to be a positive predictor for both mortality (OR=1.11, 95% CI= [1.05, 1.17], P < 0.001) and time-till death of all-cause (HR= 1.08, 95% CI= [1.04, 1.11], P < 0.001), in addition to cardiovascular-specific time-till-death (HR= 1.07, 95% CI= [1.00, 1.15], P= 0.044).

Figure 2. Kaplan-Meier plots of dichotomized mean HbA1c (A) and HbA1c variability (B) for all-cause mortality, dichotomized mean HbA1c (C) and HbA1c variability (D) for cardiovascular mortality.

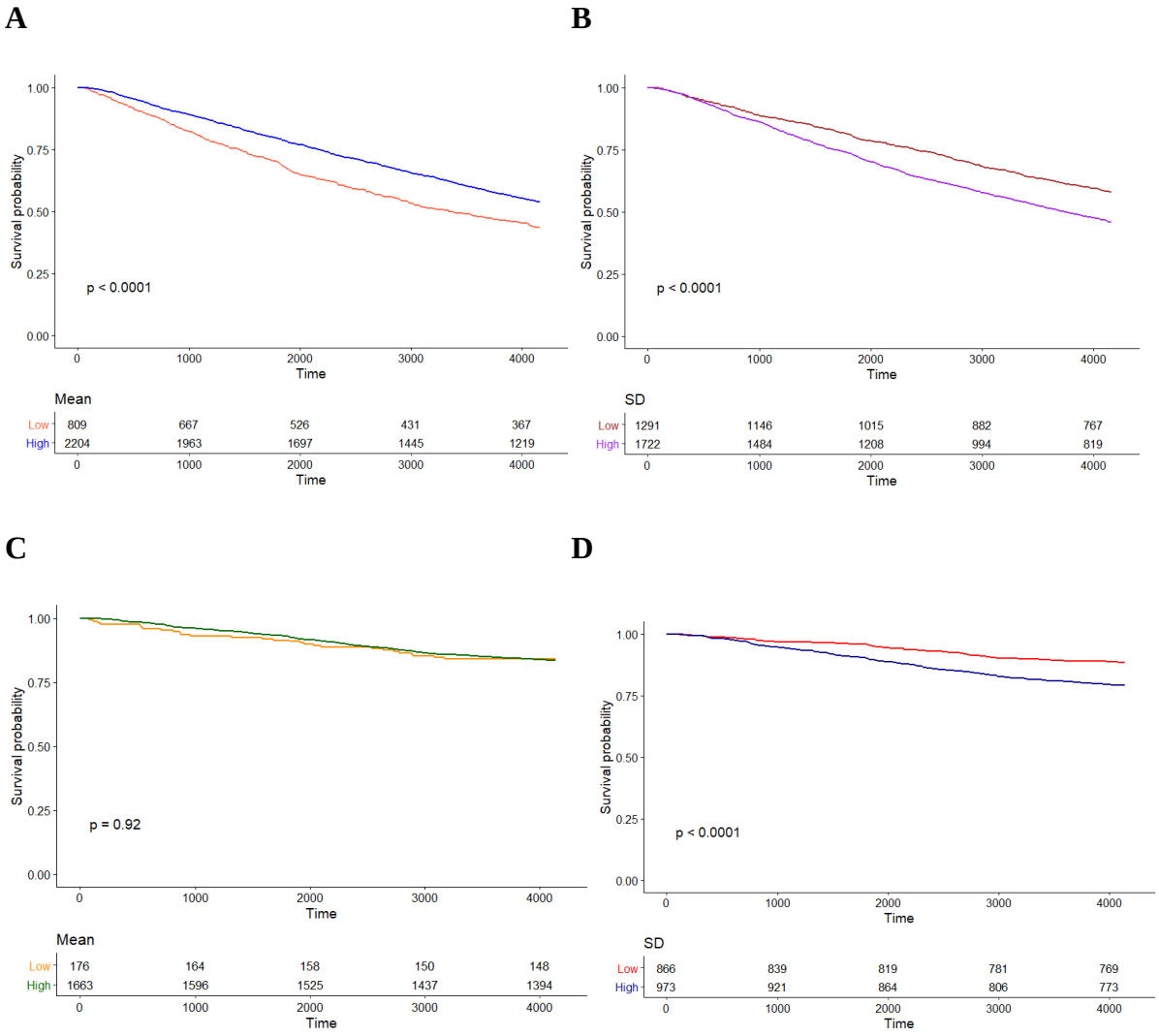


Figure A: Difference in all-cause mortality between the dichotomized groups of high (blue) vs low (orange) mean HbA1c.
Figure B: Difference in all-cause mortality between the dichotomized groups of high (pureple) vs low (brown) HbA1c variability
Figure C: Difference in cardiovascular mortality between the dichotomized groups of high (green) vs low (yellow) mean HbA1c.
Figure D: Difference in cardiovascular mortality between the dichotomized groups of high (dark blue) vs low (red) HbA1c variability.

In terms of prediction of secondary outcomes, both baseline HbA1c and HbA1c variability are significant positive predictors for the following:

1) DKA/ HHS/ coma (baseline: OR= 1.16, 95% CI= [1.04, 1.28], P= 0.006; mean: OR= 1.32, 95% CI= [1.14, 1.52], P < 0.001; SD: OR= 1.27, 95% CI= [1.02, 1.53], P= 0.018; RMS: OR= 1.29, 95% CI= [1.13, 1.47], P < 0.001; HVS: OR=1.011, 95% CI=[1.003, 1.018], P=0.004);

2) neurological diabetic complications (baseline: OR= 1.07, 95% CI= [1.00, 1.14], P= 0.033; mean: OR= 1.16, 95% CI= [1.05, 1.28], P= 0.003; SD: OR= 1.23, 95% CI= [1.07, 1.42], P= 0.004; RMS: OR= 1.15, 95% CI= [1.05, 1.27], P= 0.002; CV: OR= 1.02, 95% CI= [1.01, 1.03], P= 0.006; HVS: OR= 1.01, 95% CI= [1.00, 1.01], P=0.001);

3) ophthalmological diabetic complications (baseline: OR= 1.10, 95% CI= [1.05, 1.15], P < 0.001; mean: OR= 1.32, 95% CI= [1.23, 1.42], P < 0.001; RMS: OR= 1.28, 95% CI= [1.19, 1.37], P <0.001);

4) only mean HbA1c and RMS were predictive for microalbuminuria (mean: OR= 1.18, 95% CI= [1.08, 1.29], P < 0.001; RMS: OR= 1.16, 95% CI= [1.07, 1.27], P < 0.001) and proteinuria (mean: OR= 1.25, 95% CI= [1.14, 1.37], P < 0.001; RMS: OR=1.23, 95% CI= [1.13, 1.34], P < 0.001);

5) renal diabetic complications were predicted by mean (OR= 1.25, 95% CI= [1.03, 1.19], P = 0.006), SD (OR=1.14, 95% CI= [1.02, 1.28], P= 0.025), RMS (OR= 1.10, 95% CI= [1.03, 1.18], P= 0.007) and HVS (OR=1.005, 95% CI= [1.002, 1.008], P=0.002);

6) baseline HbA1c (OR= 1.11, 95% CI= [0.999, 1.23], P= 0.041) and HVS (OR= 1.014, 95% CI=[1.006, 1.023], P=0.001).

2.4. Discussion and limitations

2.4.1. Discussion

The major findings of the present study are that: 1) both HbA1c value and variability can predict mortality and complications in diabetics; 2) low HbA1c was associated with higher

all-cause mortality; 3) the frequency of hypoglycaemia episodes was associated with HbA1c variability; 4) hypoglycaemia frequency was predictive of both all-cause and cardiovascular mortality in diabetic patients.

In the past, a positive linear relationship is perceived between HbA1c value and all-cause mortality (69, 70). However, emerging evidence from large cohort studies suggests that there is instead an increased all-cause mortality risk at both low and high HbA1c levels (6, 60, 71, 72). In the present study, a similar U-shaped association was demonstrated for the time-till-death of both all-cause and cardiovascular mortality against HbA1c value under the generalized additive model. The increased all-cause and cardiovascular mortality risk amongst patients with low HbA1c values are demonstrated by the poorer survival of the low mean HbA1c subgroup, and the baseline HbA1c value showing hazard ratios less than unity for both mortality and time-till-death. Although the underlying mechanism between low HbA1c and increased mortality remains unclear, low HbA1c has been associated with chronic inflammation and liver function derangement (73). Furthermore, a similar U-shaped relationship has been reported for blood glucose against endothelial dysfunction and frailty (74, 75, 76). These findings, therefore, suggest that more relaxed glycaemic control for older patients with greater frailty may improve patient outcomes.

The present study provides further evidence of the predictive value of HbA1c variability. Similar to previous studies, high HbA1c variability is associated with increased risk in both all-cause and cardiovascular mortality, in addition to vascular, neurological, ophthalmological, and renal complications (8, 61, 63, 77, 78). Overall, HbA1c variability demonstrated greater predictive value for directly diabetes-caused complications in different organ systems, but its predictive value was limited for other associated conditions, such as stroke. The statistical insignificance in the prediction of cardiovascular events, contrary to existing studies, may be attributed to the single-centre nature of the study cohort, which limited

the incidence of associated conditions. Comparison between baseline, mean, SD, RMS and CV is performed to evaluate the difference in predictive performance between measurements taken upon initial visits in comparison to temporal changes, as well as the difference between different variability measures. The present findings shows that the mean and SD are more significant independent predictors. Hence, SD may be a more relevant variability measure to be used in predictive models and scores.

Although the mechanism between HbA1c variability and diabetes-induced complications remains unclear, possible explanations involving intermittent hypoglycaemia were raised. Some investigators suggested that intermittent hypoglycaemia induces the production of reactive oxygen species, and the increased oxidative stress results in endothelial dysfunction, which ultimately leads to cardiovascular complications and death (79, 80, 81, 82). Another potential theory is that hypoglycaemic episodes stimulate sympathetic activation, which stresses the cardiovascular system and affects the end-organ blood supply (12). Therefore, to elucidate the biological connections between HbA1c variability and diabetic progression, the present study examined the inter-relationship between hypoglycaemia, HbA1c variability, and diabetic outcomes. Dichotomized HbA1c variability was found to be a positive predictor of hypoglycaemic episodes, whilst hypoglycaemia itself was a positive predictor for mortality. Therefore, it can be inferred that the predictive value of HbA1c variability is at least partially contributed by the effects of intermittent hypoglycaemia.

Despite the emerging evidence for the predictive value of HbA1c variability, its clinical application remains limited by the absence of a standardized quantification method. Existing studies employed methods such as counting the frequency of significant successive differences, percentage deviation from the expected trajectory, SD, CV and HVS (60, 61, 77, 83). With different variability parameters, the present study demonstrates that the calculation of variability may be affected by the HbA1c value, and the number of measurements taken. The

resulting difference in the inherent sensitivity of the parameters results in the difference in their predictive power. Therefore, when HbA1c variability is used for prediction, different parameters are to be adopted to reduce the effect of confounding factors.

2.4.2. Limitations

Several limitations should be noted for the present study. Firstly, the size of the study cohort is limited by its single-centred nature. The resulting limitation in the incidence of diabetes-related disease events may affect the predictive power. Since free text is not captured by CDARS, this can result in the underdiagnosis of hypoglycaemic episodes and an incomplete documentation. Furthermore, like other observational studies, it is limited by the potential under-coding of comorbidities, missing data, and coding errors. The missing of data, particularly when it is not at random, can result in the introduction of bias. For example, patients with cognitive impairment and lower body mass index are more likely to have asymptomatic, or missed documentations of hypoglycaemic episodes. The underreporting of hypoglycaemic episodes amongst patients with certain risk factors can therefore lead to the underestimation of the association between hypoglycaemic and these risk factors. Additionally, the duration of diabetes, diabetic progression, and treatments prescribed were not accounted for, which can affect the interpretation of HbA1c value and variability. Given prior studies from different countries of origin reporting the association between Hba1c variability and adverse outcomes, we speculate that our findings can be generalizable to other populations. However, the external validity of the study should be confirmed in non-Chinese cohorts.

2.5. Conclusion

In conclusion, the present study demonstrated the association between high HbA1c variability and increased risk for both all-cause and cardiovascular mortality, in addition to diabetic complications across different organ systems. The association between hypoglycaemic frequency, HbA1c variability, and mortality support the hypothesis that intermittent hypoglycaemia contributes to poor outcomes in diabetic patients. Further research on larger cohorts is required to provide further evidence for the predictive value of HbA1c variability on the prognosis of diabetic patients and to shed light on its associations with hypoglycaemia.

Chapter 3. Glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning

3.1. Introduction

There is an increasing global prevalence of T2DM, with over 400 million people around the world currently suffering from the disease. (1) T2DM can lead to a variety of complications affecting the cardiovascular, neurological, renal and other systems, placing significant burdens on healthcare systems globally. (84, 85, 86) Given the ageing population, an increasing proportion of diabetic patients are elderly with multiple comorbidities, leading to a call for a more personalised and patient-centred approach in diabetic management over recent years. (87, 88, 89) This raises the need for new parameters for monitoring diabetes, other than blood glucose, to improve the sensitivity towards the disease progression across different organ systems. (90, 91, 92, 93, 94) Diabetic patients who are on insulin are more advanced in the disease life course, and as such are at a higher risk of complications and death.

Recently, HbA1c and lipid variability have attracted attention in their potential use for diabetic monitoring and risk stratification for adverse outcomes. However, existing studies focused on cardiovascular events and mortality. (57, 95, 96) Although the exact pathways of pathogenesis by HbA1c and different lipid variability are unclear and appear to be divergent, the resulting chronic inflammation and endothelial dysfunction may have led to the presentation of systemic complications in diabetes. (77, 97, 98) Others suggest that raised variability in biomarkers reflects lifestyle changes, incomplete treatment adherence, pharmacotherapy prescribed, and generalized frailty. (99, 100, 101) Random survival forest (RSF) is a class of machine learning algorithms for survival analysis (102). Although RSF has been previously used to predict for protein interactions, and mortality amongst patients with HF, it has not been used in the prediction of cardiovascular risks amongst patients with T2DM before (103, 104). The advantage of RSF is that it can reduce the variance and bias within the input variables and automatically consider nonlinear effects and high-level interactions among

these variables. Thus, RSF can be applied to select and rank variables based on their importance. In this study, we aim to evaluate the predictive value of glycaemic and lipid variability towards a wide range of adverse outcomes in diabetes and that risk prediction is more accurate using RSF. The present study is the first to use RSF in the prediction of cardiovascular adverse events amongst patients with T2DM.

3.2. Methods

3.2.1. Study population

The present study is a territory-wide observational study that collects data from 43 public hospitals in Hong Kong. The study was approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee. It was performed in accordance with the Declaration of Helsinki as well as relevant guidelines and regulations. The cohort consists of diabetic patients who have been prescribed insulin from outpatient clinics of any public hospitals managed by the Hong Kong Hospital Authority between January 1st, 2009, to December 31st, 2009. Patients were not required to be on insulin for a minimum period. Through CDARS, the cohort was identified, and the data was extracted. The system has been utilized for epidemiological research by multiple research teams, including our team, in the past (68, 105, 106, 107).

3.2.2. Patient data

Clinical outcomes, patient characteristics and pharmacological treatment details were extracted. The patient outcomes from January 1st, 2009 to December 31st, 2019 were extracted. Patients were followed up from January 1st, 2009 to either death or December 31st, 2019. The primary outcome is all-cause mortality, and the secondary outcomes, as defined by their ICD-

9 codes include 1) neurological, ophthalmological and renal diabetic complications, 2) dementia, 3) osteoporosis, 4) peripheral vascular disease (PVD), 5) intracranial haemorrhage (ICH), 6) ischemic stroke and TIA, 7) IHD, AMI and HF, 8) AF.

The extracted patient parameters are as follows**.** The duration of diabetes at baseline was extracted based on the following three criteria, selected based on whichever is earlier: 1) earliest ICD-9 coding of diabetes mellitus; 2) earliest HbA1c > 6.5mmol/L; 3) earliest fasting blood glucose > 7mmol/L. The mean daily dose of anti-diabetic and cardiovascular medications drug classes was reported. The mean daily dose is derived by multiplying the daily dose frequency against the drug dose, then averaged by all patients that were prescribed drugs of the specific drug class. In terms of biochemical data, the baseline neutrophil-lymphocyte ratio (NLR) was derived by dividing the baseline absolute neutrophil count by the lymphocyte count. To assess glycaemic and lipid variability, data for the following variables between January 1st, 2004 and December 31st, 2008 were obtained: 1) HbA1c, 2) total cholesterol, 3) HDL-C, 4) LDL-C, 5) total triglyceride. LDL-C includes both findings from direct and calculated measurements. Furthermore, the frequency of hypoglycaemic episodes across the entire follow-up period from laboratory tests taken during outpatient, inpatient and accident and emergency settings was extracted. Each episode is defined by random or FBG < 3.9 mg/mmol. Additionally, the presence of anaemia, defined by haemoglobin < 13 g/dL and <12g/dL for male and female patients respectively, was extracted. The presence of iron deficiency, defined by ferritin < 67.4 pmol/L, was also extracted. Only patients with three or more measurements for the specific parameter were included in the variability analysis of the respective parameter.

3.2.3. Statistical analysis

Temporal variability was examined using the derivation of SD and CV. CV was given by the temporal SD divided by the temporal mean, then multiplied by 100. Univariate Cox

regression was applied to identify significant predictors from demographic variables, biochemical parameters, and anti-diabetic agents prescribed for the various adverse outcomes. GLP-1A (n=13) and meglitinide (n=9) were excluded from the analysis due to the limited number of patients prescribed the drugs. The HR and 95% CI were presented for each predictor. Patients with missing data were excluded from the analysis for that variable. Predictors with P-value < 0.10 under univariate analysis for all-cause mortality are then selected to undergo multivariate Cox regression. Patients were excluded from the multivariate analysis if they did not have at least three measurements for the assessment of variability, or if there were missing data in any of the significant predictors found under univariate Cox analysis.

To examine the inter-relationship between HbA1c variability, intermittent hypoglycaemia, and chronic inflammation, Gaussian, and Poisson regression were used to assess the correlations of HbA1c variability against baseline NLR and hypoglycaemia frequency respectively. Gaussian regression is a non-parametric method to assess the association between two continuous variables, hence suitable to assess the inter-relationship between HbA1c/ lipid variability and baseline NLR. Poisson regression is a model that allows the assessment between a count variable, in this case, hypoglycaemic frequency, and continuous variables, such as HbA1c/ lipid variability. Gaussian regression was also used to assess the association between the lipid parameters, and lipid indices against baseline NLR. The OR is reported for both Poisson and Gaussian regression. Statistical significance is defined as P-value <0.05. Statistical analyses were performed using RStudio software (Version: 1.1.456) and Python (Version: 3.6).

3.2.4. Development of a regularized and weighted random survival forest model

RSF is a machine-learning modelling technique that can capture complex survival data structures and overcome the restrictive assumption of the Cox proportional model to better

uncover the nonlinear relationships between covariates and the time of event outcome. (108) In contrast, assumptions about specialized basis functions in Cox models are not efficient for assessing the nonlinear effects by transformations or expanding the design matrix. The RSF model is constructed with an ensemble tree method for analysis of right-censored survival data, extended from Breiman's random forest. It is an efficient ensemble learning method by injecting randomization into base learning processes and has become one of the most efficient models in survival analysis.

In this study, the time for RSF survival learning is defined as the duration from baseline date to event presentation or mortality/ study end date if no event presentation before mortality and study end. As shown in **Figure 3** for the workflow of the regularized and weighted RSF model, the regularized and weighted RSF model estimates the forest survival function through a tree ensemble approach. Usually, the ensemble assigns equal weights to different survival decision trees in a RSF. In the present study, different weights were assigned to different survival trees to account for the heterogeneity between the ensembled decision tries. The assigned weights were learned to minimize the overall loss function (e.g., the log-likelihood we used in this study). To reduce overfitting, we adopt an L2 regularization strategy with an optimal regularization strength parameter for the log-likelihood loss function in the model. The regularization parameters were determined by using 80% of the patients in the cohort under a five-fold cross-validation.

With the RSF model, the learning results can be interpreted by estimating the relative importance and minimal depth of the individual variables. The importance value for the variable of interest is the prediction error for the original ensemble event-specific cumulative probability function, excluding the out-of-bag instances, subtracted from the prediction error for the new ensemble obtained using randomizing assignments of the variable (109, 110). The prediction errors are computed using squared loss. A larger importance value indicates a higher

predictive strength, whereas importance values equal or less than renders the variable nonpredictive. The minimal depth approach (111) identifies variables of averages the depth of the first split for each variable over all trees within the final forest to identify the most frequently split nodes nearest to the root, which partition the largest samples, and therefore has the greatest impact on the prediction on the clinical outcome.

Significant variables from the univariate Cox regression were used as inputs into the regularized and weighted RSF model. The performance of regularized and weighted RSF, RSF and the Cox model are compared. Missing values are "-1" padded. The machine learning models are trained on the training set using five-fold cross validation. The model's discrimination performance is accessed by Harrell's C-index, which is a generalization of the AUC that can handle right-censored data, to estimate the efficiency of the model at ranking survival times.

3.3. Results

3.3.1. Clinical and biochemical characteristics

The study cohort consists of 25 186 patients (mean age= 63.0, interquartile range [IQR] of age= 15.1 years, male = 50.4%, type 1 diabetes mellitus= 7.37%, baseline diabetes duration = 2.84 ± 2.54 years, total duration= 69332 patient-years, daily insulin dosage: 20.2 ± 12.6 units). Patients with type 1 DM were included, despite the primary focus of the study to be around T2DM, is to maximize the cohort size available for the evaluation of machine learning models. A graphical illustration of the methodology is shown in **Figure 3**. **Tables 4 and 5** display the discrete and continuous baseline characteristics of the study cohort respectively. It should be noted that there is a large SD in the continuous baseline characteristics shown on **Table 5,** The most prevalent pre-existing comorbidity is hypertension (35.6%), followed by

ophthalmological conditions (32.2%), and IHD (16.2%). The mean daily insulin dosage regimen of the cohort is 20.2 ± 12.6 units per day. Other classes of anti-diabetic agents prescribed include: 1) biguanide (n= 14 522, mean daily dose= 1682 ± 882 mg/day), 2) sulphonylurea (n= 10 459, mean daily dose= 191 ± 312mg/day), 3) thiazolidinedione (n= 890, mean daily dose= 7.59 ± 9.03mg/day), 4) alpha-glucosidase inhibitor (n= 751, mean daily dose= 138 ± 128mg/day), 5) DPP4I (n= 113, mean daily dose= 91.6 ± 27.5mg), 6) GLP-1A (n= 13, mean daily dose= 11.9 ± 3.96mg/day), 7) meglitinide (n= 9, mean daily dose= 1.17 ± 0.887mg/day). The following classes of cardiovascular medications were prescribed: 1) ACEI/ ARB (n= 15 059, mean daily dose= 17.4 ± 37.0mg/day), 2) CCB (n= 10 986, mean daily dose= 60.0 ± 50.6mg/day), 3) lipid-lowering agents (n= 10 685, mean daily dose= 23.3 ± 68.4 mg/day), 4) aspirin (n= 9114, mean daily dose= 102 ± 54.1 mg/day), 5) diuretic (n= 7 349, mean daily dose= 40.3 ± 64.2mg/day), 6) beta-adrenergic receptor blocker (n= 7 082, mean daily dose= 79.4 ± 81.6mg/day).

Figure 3. Workflow of regularized and weighted random survival forest model

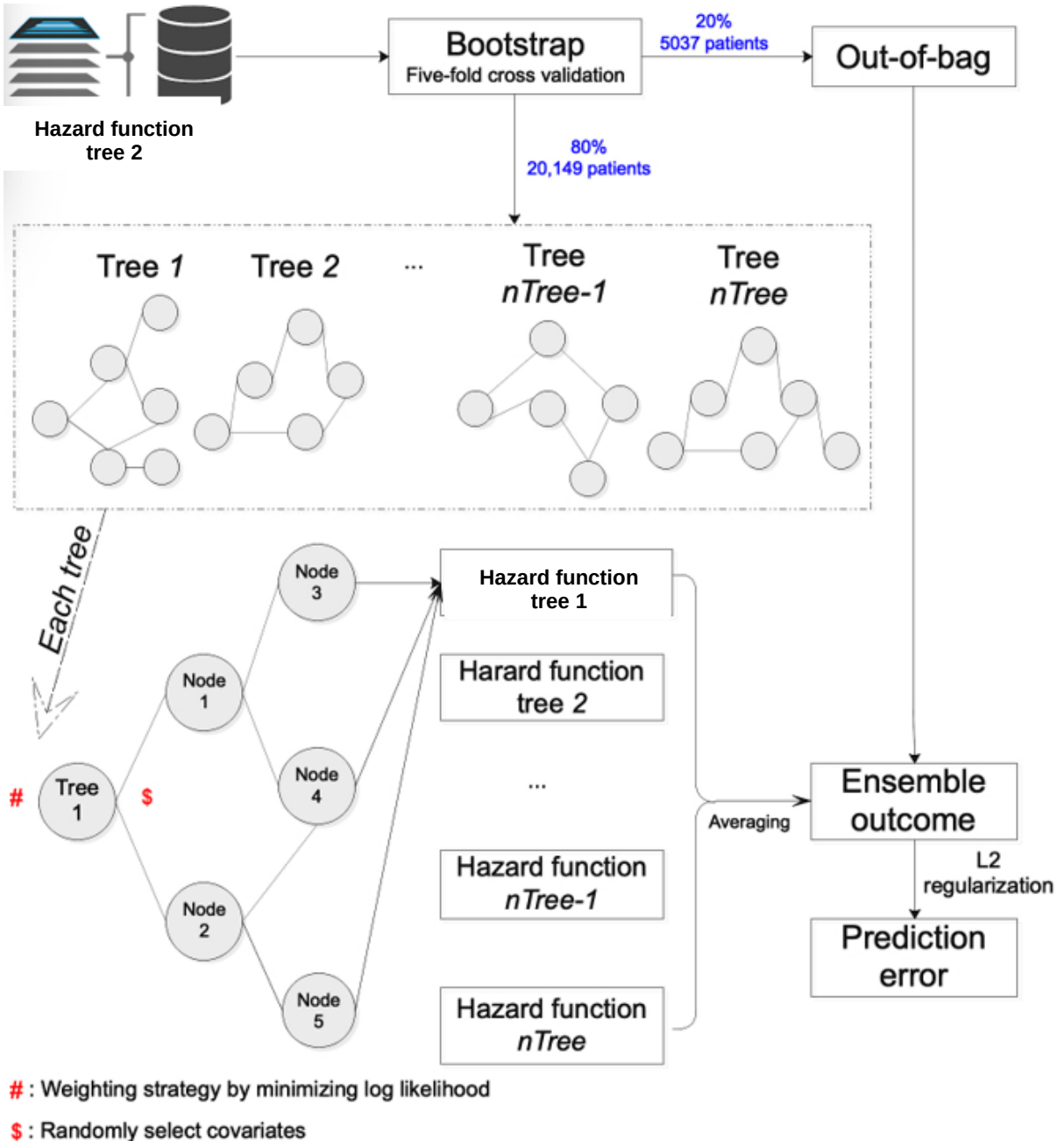


Figure 3 shows data were bootstrapped to form multiple decision trees, cross-validated using trees with out-of-bag instances. Within each tree, each node are predictors to ultimately lead to the ensemble outcome. Prediction error is calculated under L2 regularization.

Table 4. Discrete baseline characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning

	Patient Percentage (%)	Number of Patients
Demographic		
Male	50.4	12694
Type 1 Diabetes Mellitus	7.37	1856

Comorbidities		
Hypertension	35.6	8966
Ophthalmological Complications	32.2	8110
Ischemic Heart Disease	16.2	4080
Ischemic Stroke and Transient Ischemic Attack	11.8	2972
Heart Failure	9.8	2468
Chronic Renal Disease	8.8	2216
Chronic Liver Disease	5.8	1461
Acute Myocardial Infarction	5.1	1284
Chronic Obstructive Pulmonary Disease	3.5	882
Anti-diabetic Medication		
Biguanide	57.6	14522
Sulphonylurea	41.5	10459
Thiazolidinedione	3.5	890
Alpha-Glucosidase	3.0	751
Dipeptidyl Peptidase-4 Inhibitor	0.4	113
Glucagon-Like Peptide-1 Receptor Agonist	<0.1	13
Cardiovascular medication		
Angiotensinogen-Converting Enzyme Inhibitor/ Angiotensin-Receptor Blocker	59.8	15059
Calcium Channel Blocker	43.6	10986
Lipid-Lowering Agents	42.4	10685
Aspirin	36.2	9114
Diuretic	29.2	7349
Beta-Adrenergic Receptor Blocker	28.1	7082

Table 5. Continuous baseline characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning

	Mean	Standard Deviation
Urinalysis		
Albumin/Creatinine Ratio (mg/mmol)	38.1	121
Creatinine Clearance (ml/min)	54.1	35.9
Spot Protein (g/d)	1.17	1.96
Spot Albumin (mg/L)	170	545
Spot Glucose (mmol/L)	12.5	6.68
24-hours Total Protein (g/d)	1.17	1.97
24-hours Total Albumin (mg/d)	271	695
Baseline Blood Test		
Fasting Glucose (mmol/L)	8.96	3.75
Random Glucose (mmol/L)	12.3	7.47
HbA1c (%)	8.56	1.94
Total Cholesterol (mmol/L)	4.74	1.12

High Density Lipoprotein (HDL) Cholesterol (mmol/L)	1.24	0.403
Calculated Low Density Lipoprotein (LDL) Cholesterol (mmol/L)	2.74	0.927
Direct LDL Cholesterol (mmol/L)	2.80	0.925
Triglyceride (mmol/L)	1.80	1.72
Renal Function Test		
Creatinine (umol/L)	144	159
Sodium (mmol/L)	139	3.33
Potassium (mmol/L)	4.31	0.506
Urate (umol/L)	0.408	0.129
Urea (mmol/L)	8.82	6.04
Liver Function Test		
Albumin (g/L)	39.2	5.56
Alanine Aminotransferase (ALT) (U/L)	24.3	21.6
Alkaline Phosphatase (ALP) (U/L)	85.2	47.0
Total Bilirubin (umol/L)	11.3	8.98
Total Protein (g/L)	74.4	7.13
Complete Blood Count		
Haemoglobin (g/dL)	12.5	1.99
Mean Corpuscular Haemoglobin (MCH) (pg)	29.7	2.95
Mean Corpuscular Haemoglobin Concentration (MCHC) (g/dL)	34.0	0.952
Mean Corpuscular Volume (MCV) (fL)	87.2	7.44
Hematocrit (L/L)	0.376	0.539
Basophil ($x10^9$/L)	0.029	0.042
Eosinophil ($x10^9$/L)	0.223	0.235
Lymphocyte ($x10^9$/L)	1.87	0.867
Monocyte ($x10^9$/L)	0.538	0.266
Neutrophil ($x10^9$/L)	5.47	2.79
Platelet ($x10^9$/L)	256	83.3
Red Blood Cell ($x10^{12}$/L)	4.26	0.740
White Blood Cell ($x10^9$/L)	8.09	2.91

3.3.2. Anti-diabetic drug classes and outcomes

Different classes of anti-diabetic agents are associated with adverse outcomes differently. Thiazolidinedione lowers the risk of neurological complications (HR= 0.718, 95% CI= [0.539, 0.956], p= 0.023) and HF (HR= 0.72, 95% CI= [0.54, 0.96], p < 0.0001), whilst biguanide only lowers the risk of HF (HR= 0.62, 95% CI= [0.56, 0.68], p < 0.0001). The risk for adverse cardiovascular events was raised by sulphonylurea, biguanide, and alpha-glucosidase inhibitors. Sulphonylurea is associated with an increased risk of renal complications (HR= 1.29, 95% CI= [1.22, 1.36], p < 0.0001) and dementia (HR= 1.22, 95%

CI= [1.08, 1.39], p= 0.002), whilst biguanide is related to ophthalmological complications (HR= 1.09, 95% CI= [0.937, 1.26], p < 0.0001).

3.3.3. Adverse outcome and predictors

The characteristics of the adverse outcomes and biochemical predictors are detailed in **Tables 6** and **7** respectively. Anaemia occurred in 39.1% (n= 9848) of the cohort, with iron deficiency presented in 9.76% of the 2100 patients with ferritin measured. Throughout the study period, 12 372 incidences of death took place (male= 52.6%, age of death= 69.7 ± 12.0). The most common adverse outcomes were death (49.1%), renal (21.4%), and ophthalmological complications (18.7%). Ophthalmological (onset age= 62.8 ± 11.9), neurological (onset age= 64.2 ± 11.9) and renal diabetic complications (onset age= 66.5 ± 12.2) had the earliest onset, whilst osteoporosis (onset age= 72.1 ± 11.3) and dementia (onset age= 74.4 ± 8.30) occurred latest on average, patients in the present cohort experience 1.74 ± 1.72 adverse outcomes.

Table 6. Adverse outcome characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning

Outcome	Number of events	Incidence rate	Age of onset	Number of comorbidities
Mortality	12372	49.12%	69.7 ± 12.0	2.71 ± 1.66
Renal	5389	21.40%	66.5 ± 12.2	3.51 ± 1.67
Ophthalmological	4705	18.68%	62.8 ± 11.9	3.11 ± 1.81
Ischemic Heart Disease	4532	17.99%	66.8 ± 11.6	3.70 ± 1.74
Acute myocardial infarction	3178	12.62%	68.3 ± 11.1	4.15 ± 1.59
Neurological	1861	7.39%	64.2 ± 11.9	4.03 ± 1.77
Atrial Fibrillation	1846	7.33%	70.4 ± 10.3	3.75 ± 1.74
Heart Failure	1810	7.19%	68.9 ± 11.4	4.61 ± 1.45
Ischemic Stroke	1350	5.36%	69.2 ± 10.9	3.55 ± 1.79
Intracranial Hemorrhage	1049	4.17%	68.4 ± 11.7	3.45 ± 1.67
Dementia	952	3.78%	74.4 ± 8.30	3.37 ± 1.68
Peripheral vascular disease	711	2.82%	66.6 ± 12.4	4.39 ± 1.85
Osteoporosis	275	1.09%	72.1 ± 11.3	3.01 ± 1.68

Table 7. Biochemical predictor characteristics of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning

Predictors	Mean	Standard Deviation
HbA1c		
Baseline (%, n=24064)	8.56	1.94
Mean (%, n=22625)	8.64	1.36
Standard Deviation	1.28	0.851
Coefficient of Variation	14.5	8.76
Total Cholesterol (TC)		
Baseline (mmol/L, n=23532)	4.74	1.12
Mean (mmol/L, n=20445)	4.82	0.871
Standard Deviation	0.663	0.459
Coefficient of Variation	13.5	7.95
High Density Lipoprotein-Cholesterol (HDL-C)		
Baseline (mmol/L, n=23178)	1.24	0.402
Mean (mmol/L, n=19303)	1.25	0.362
Standard Deviation	0.161	0.100
Coefficient of Variation	1.24	0.403
Low Density Lipoprotein-Cholesterol (LDL-C)		

Baseline (mmol/L, n=23075)	1.24	0.913
Mean (mmol/L, n=18803)	2.78	0.734
Standard Deviation	0.553	0.359
Coefficient of Variation	20.3	12.5
Triglyceride		
Baseline (mmol/L, n=23518)	1.80	1.72
Mean (mmol/L, n=20398)	1.86	1.43
Standard Deviation	6.90	1.15
Coefficient of Variation	30.8	17.8
Other Tests		
Baseline NLR	3.80	4.16
Baseline Haemoglobin Count (g/dL)	12.5	1.99
Hypoglycaemia Frequency	0.537	1.38

The number of patients included for the calculation of the mean is the same as the number of patients included for the calculation of standard deviation and coefficient of variation.

Multivariate Cox regression analysis was applied to 7 913 patients from the study cohort. The multivariate Cox regression for all-cause mortality is presented in **Table 8**. Mean HbA1c was found to be protective against mortality in univariate analysis (HR= 0.964, $p < 0.0001$), but became predictive in multivariate analysis. However, after adjusting for haematological malignancies, iron deficiency status and lipid-lowering drug use (n=652), HbA1c mean and variability did not remain significant predictors. Amongst the lipid predictors (n=7913), only HDL-C mean (HR= 0.60, 95% CI= [0.51, 0.71], $p < 0.0001$) and SD (HR=2.18, 95% CI= [1.51, 3.14], $p < 0.0001$) remained significant after adjusting for cancer status and lipid-lowering agent use.

Table 8. Multivariate Cox regression of all-cause mortality in diabetes mellitus using machine learning

Predictor	Hazard Ratio (HR)	95% Confidence Interval (CI)	P-Value
Age	1.04	[1.03, 1.04]	**<0.0001**
Male	1.18	[1.11, 1.27]	**<0.0001**
Diabetes Duration	0.956	[0.943, 0.970]	**<0.0001**
HbA1c			
Mean	1.09	[1.04, 1.15]	**<0.001**
Standard Deviation	1.10	[0.825, 1.47]	0.511
Coefficient of Variation	0.998	[0.973, 1.02]	0.869

Total Cholesterol (TC)			
Mean	1.14	[0.994, 1.30]	0.061
Standard Deviation	0.787	[0.501, 1.24]	0.299
Coefficient of Variation	1.02	[1.00, 1.05]	**0.050**
High Density Lipoprotein-Cholesterol (HDL-C)			
Mean	0.603	[0.513, 0.708]	**<0.0001**
Standard Deviation	2.19	[1.52, 3.14]	**<0.0001**
Low Density Lipoprotein-Cholesterol			
Mean	0.916	[0.811, 1.03]	0.157
Standard Deviation	1.19	[0.866, 1.64]	0.281
Coefficient of Variation	0.992	[0.983, 1.00]	0.062
Triglyceride (TG)			
Baseline	0.996	[0.979, 1.01]	0.694
Mean	1.06	[0.993, 1.14]	0.080
Standard Deviation	0.932	[0.851, 1.02]	0.126
Coefficient of Variation	0.998	[0.995, 1.00]	0.190
Other Tests			
Baseline Neutrophil-Lymphocyte Ratio	1.01	[1.01, 1.02]	**<0.001**
Baseline Haemoglobin Count	0.911	[0.889, 0.934]	**<0.0001**
Baseline Anaemia	1.08	[0.981, 1.19]	0.119
Hypoglycaemia Frequency	1.03	[1.01, 1.05]	**0.002**
Anti-Diabetic Agent			
Sulphonylurea	1.08	[1.02, 1.16]	**0.015**
Biguanide	0.616	[0.575, 0.660]	**<0.0001**
Dipeptidyl peptidase-4 Inhibitor	0.706	[0.424, 1.18]	0.181
Thiazolidinedione	0.885	[0.761, 1.03]	0.110

In terms of prediction of secondary outcomes, the predictors were similar to those for all-cause mortality and are summarized in **Table 9**. HbA1c variability is predictive of adverse outcomes besides osteoporosis, ischemic stroke, and AMI. HbA1c CV is mildly protective of IHD (HR= 0.996, 95% CI= [0.993, 1.00], p = 0.046). In terms of lipid predictors, elevated mean total cholesterol is predictive of most adverse outcomes, except for AF (HR= 0.889, 95% CI= [0.838, 0.943], p < 0.0001). Increased mean HDL-C lowers the risk for adverse outcomes, except for osteoporosis (HR= 1.78, 95% CI= [1.29, 2.44], p < 0.001). Heterogenous predictions were noted for HDL-C variability and mean LDL-C. By contrast, increased LDL-C variability

predicts an increased risk for various adverse outcomes. In terms of the predictiveness of triglyceride level, both its value and variability were found to be predictive of different adverse outcomes, except for CV of triglyceride being protective against osteoporosis (HR= 0.990, 95% CI= [0.981, 0.998], p= 0.020). Baseline NLR and frequency of hypoglycaemic episodes were predictive for a similar set of adverse outcomes, where they increased the risk for PVD), HF, and all-cause mortality, but were associated with a lower risk for ophthalmological complications.

Table 9. Univariate Cox regression for adverse outcomes of glycaemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning

Predictor	Hazard Ratio (HR)	95% Confidence Interval (CI)	P-Value
Neurological Complications (n=1861)			
Age	1.01	[1.00, 1.01]	**<0.001**
Male	1.47	[1.34, 1.62]	**<0.0001**
Diabetes Duration	1.04	[1.03, 1.06]	**<0.0001**
HbA1c			
Baseline	0.999	[0.974, 1.02]	0.934
Mean	1.18	[1.14, 1.22]	**<0.0001**
Standard Deviation	1.13	[1.07, 1.19]	**<0.0001**
Coefficient of Variation	1.01	[1.00, 1.01]	**0.014**
Total Cholesterol			
Baseline	0.960	[0.917, 1.01]	0.081
Mean	1.06	[1.00, 1.12]	**0.042**
Standard Deviation	1.16	[1.06, 1.27]	**0.001**
Coefficient of Variation	1.01	[1.00, 1.02]	**0.001**
HDL Cholesterol			
Baseline	1.01	[0.891, 1.15]	0.863
Mean	0.634	[0.544, 0.741]	**<0.0001**
Standard Deviation	0.555	[0.323, 0.954]	**0.033**
Coefficient of Variation	1.00	[0.994, 1.01]	0.881
LDL Cholesterol			
Baseline	1.00	[0.946, 1.06]	0.964
Mean	1.03	[0.961, 1.11]	0.398
Standard Deviation	1.27	[1.11, 1.45]	**<0.001**
Coefficient of Variation	1.01	[1.00, 1.01]	**0.001**
Triglyceride			
Baseline	0.962	[0.928, 0.997]	**0.035**
Mean	1.07	[1.05, 1.10]	**<0.0001**
Standard Deviation	1.05	[1.02, 1.09]	**0.002**

Coefficient of Variation	1.00	[0.999, 1.00]	0.229
Baseline NLR	0.994	[0.981, 1.01]	0.357
Baseline Haemoglobin Count	0.945	[0.923, 0.968]	<**0.0001**
Baseline Anaemia	1.17	[1.06, 1.29]	**0.002**
Hypoglycaemia Frequency	1.01	[0.977, 1.04]	0.612
Anti-Diabetic Agent			
Sulphonylurea	0.936	[0.853, 1.03]	0.162
Biguanide	0.978	[0.892, 1.07]	0.629
DPP4 Inhibitor	0.923	[0.392, 1.73]	0.608
Thiazolidinedione	0.718	[0.539, 0.956]	**0.023**
Alpha-Glucosidase Inhibitor	1.05	[0.812, 1.37]	0.694
Ophthalmological Complications (n=4705)			
Age	0.999	[0.997, 1.00]	0.427
Male	1.01	[0.955, 1.07]	0.694
Diabetes Duration	1.11	[1.10, 1.12]	<**0.0001**
HbA1c			
Baseline	1.01	[0.997, 1.03]	0.109
Mean	1.25	[1.23, 1.28]	<**0.0001**
Standard Deviation	1.07	[1.04, 1.11]	<**0.0001**
Coefficient of Variation	0.999	[0.995, 1.00]	0.508
Total Cholesterol			
Baseline	0.996	[0.969, 1.03]	0.804
Mean	1.10	[1.06, 1.14]	<**0.0001**
Standard Deviation	1.06	[0.994, 1.13]	0.077
Coefficient of Variation	1.00	[0.997, 1.01]	0.760
HDL Cholesterol			
Baseline	0.993	[0.917, 1.08]	0.870
Mean	0.842	[0.768, 0.923]	<**0.001**
Standard Deviation	0.432	[0.305, 0.613]	<**0.0001**
Coefficient of Variation	0.990	[0.986, 0.995]	<**0.0001**
LDL Cholesterol			
Baseline	0.994	[0.959, 1.03]	0.720
Mean	1.05	[1.00, 1.10]	**0.032**
Standard Deviation	1.06	[0.974, 1.16]	0.170
Coefficient of Variation	1.00	[0.999, 1.00]	0.342
Triglyceride			
Baseline	1.01	[0.993, 1.03]	0.245
Mean	1.07	[1.05, 1.08]	<**0.0001**
Standard Deviation	1.05	[1.03, 1.07]	<**0.0001**
Coefficient of Variation	0.999	[0.997, 1.00]	0.287
Baseline NLR	0.986	[0.977, 0.995]	**0.003**
Baseline Haemoglobin Count	1.01	[0.993, 1.03]	0.265
Baseline Anaemia	0.939	[0.881, 1.00]	0.051

Hypoglycaemia Frequency	0.949	[0.926, 0.972]	**<0.0001**
Anti-Diabetic Agent			
Sulphonylurea	0.972	[0.918, 1.03]	0.344
Biguanide	1.41	[1.33, 1.50]	**<0.0001**
DPP4 Inhibitor	1.01	[0.660, 1.56]	0.954
Thiazolidinedione	1.09	[0.937, 1.26]	0.271
Alpha-Glucosidase Inhibitor	0.948	[0.798, 1.13]	0.540
Renal Complications (n=5389)			
Age	1.02	[1.02, 1.02]	**<0.0001**
Male	1.12	[1.06, 1.18]	**<0.0001**
Diabetes Duration	1.05	[1.04, 1.06]	**<0.0001**
HbA1c			
Baseline	1.00	[0.987, 1.02]	0.813
Mean	1.13	[1.11, 1.15]	**<0.0001**
Standard Deviation	1.07	[1.04, 1.10]	**<0.0001**
Coefficient of Variation	1.00	[1.00, 1.01]	0.089
Total Cholesterol			
Baseline	0.999	[0.973, 1.03]	0.943
Mean	1.08	[1.05, 1.12]	**<0.0001**
Standard Deviation	1.18	[1.11, 1.24]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.01]	**<0.0001**
HDL Cholesterol			
Baseline	1.00	[0.929, 1.08]	0.979
Mean	0.510	[0.464, 0.560]	**<0.0001**
Standard Deviation	0.372	[0.268, 0.516]	**<0.0001**
Coefficient of Variation	1.00	[0.996, 1.00]	0.906
LDL Cholesterol			
Baseline	1.00	[0.969, 1.04]	0.942
Mean	1.04	[0.998, 1.09]	0.062
Standard Deviation	1.27	[1.18, 1.38]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.01]	**<0.0001**
Triglyceride			
Baseline	0.989	[0.970, 1.01]	0.231
Mean	1.10	[1.08, 1.11]	**<0.0001**
Standard Deviation	1.06	[1.04, 1.08]	**<0.0001**
Coefficient of Variation	1.00	[1.00, 1.00]	**0.045**
Baseline NLR	1.00	[0.994, 1.01]	0.821
Baseline Haemoglobin Count	0.908	[0.896, 0.921]	**<0.0001**
Baseline Anaemia	1.46	[1.38, 1.55]	**<0.0001**
Hypoglycaemia Frequency	1.00	[0.982, 1.02]	0.865
Anti-Diabetic Agent			
Sulphonylurea	1.29	[1.22, 1.36]	**<0.0001**
Biguanide	0.970	[0.920, 1.02]	0.275

DPP4 Inhibitor	0.761	[0.485, 1.19]	0.234
Thiazolidinedione	1.00	[0.867, 1.16]	0.986
Alpha-Glucosidase Inhibitor	1.14	[0.986, 1.33]	0.075
Dementia (n=952)			
Age	1.08	[1.07, 1.08]	**<0.0001**
Male	0.717	[0.630, 0.815]	**<0.0001**
Diabetes Duration	0.981	[0.956, 1.01]	0.140
HbA1c			
Baseline	1.00	[0.968, 1.04]	0.891
Mean	1.00	[0.953, 1.05]	0.963
Standard Deviation	1.11	[1.03, 1.19]	**0.005**
Coefficient of Variation	1.01	[1.00, 1.02]	**0.003**
Total Cholesterol			
Baseline	0.984	[0.922, 1.05]	0.617
Mean	0.945	[0.871, 1.03]	0.181
Standard Deviation	1.01	[0.864, 1.17]	0.935
Coefficient of Variation	1.00	[0.995, 1.01]	0.411
HDL Cholesterol			
Baseline	1.02	[0.852, 1.23]	0.818
Mean	0.956	[0.779, 1.17]	0.664
Standard Deviation	1.35	[0.660, 2.74]	0.414
Coefficient of Variation	1.01	[0.995, 1.02]	0.342
LDL Cholesterol			
Baseline	1.05	[0.971, 1.14]	0.220
Mean	0.925	[0.835, 1.03]	0.140
Standard Deviation	1.17	[0.960, 1.42]	0.119
Coefficient of Variation	1.01	[1.00, 1.01]	**0.013**
Triglyceride			
Baseline	0.990	[0.946, 1.04]	0.643
Mean	1.00	[0.956, 1.05]	0.877
Standard Deviation	0.981	[0.917, 1.05]	0.569
Coefficient of Variation	0.999	[0.995, 1.00]	0.532
Baseline NLR	1.01	[0.993, 1.02]	0.302
Baseline Haemoglobin Count	0.927	[0.897, 0.957]	**<0.0001**
Baseline Anaemia	1.43	[1.25, 1.64]	**<0.0001**
Hypoglycaemia Frequency	1.01	[0.964, 1.05]	0.723
Anti-Diabetic Agent			
Sulphonylurea	1.22	[1.08, 1.39]	**0.002**
Biguanide	0.962	[0.846, 1.09]	0.548
DPP4 Inhibitor	0.459	[0.115, 1.84]	0.271
Thiazolidinedione	0.857	[0.592, 1.24]	0.414
Alpha-Glucosidase Inhibitor	0.729	[0.473, 1.12]	0.152
Osteoporosis (n=275)			

Age	1.06	[1.04, 1.07]	**<0.0001**
Male	0.249	[0.186, 0.335]	**<0.0001**
Diabetes Duration	1.01	[0.961, 1.06]	0.776
HbA1c			
Baseline	0.957	[0.893, 1.03]	0.205
Mean	1.00	[0.916, 1.10]	0.943
Standard Deviation	0.879	[0.747, 1.04]	0.123
Coefficient of Variation	0.985	[0.969, 1.00]	0.059
Total Cholesterol			
Baseline	0.997	[0.885, 1.12]	0.954
Mean	1.06	[0.918, 1.23]	0.411
Standard Deviation	0.966	[0.720, 1.30]	0.819
Coefficient of Variation	0.997	[0.981, 1.01]	0.757
HDL Cholesterol			
Baseline	1.01	[0.723, 1.41]	0.958
Mean	1.78	[1.29, 2.44]	**<0.001**
Standard Deviation	1.49	[0.412, 5.42]	0.542
Coefficient of Variation	0.987	[0.968, 1.01]	0.222
LDL Cholesterol			
Baseline	1.04	[0.902, 1.21]	0.566
Mean	1.03	[0.854, 1.24]	0.773
Standard Deviation	0.940	[0.639, 1.38]	0.752
Coefficient of Variation	0.996	[0.985, 1.01]	0.489
Triglyceride			
Baseline	1.04	[0.991, 1.09]	0.117
Mean	0.918	[0.818, 1.03]	0.145
Standard Deviation	0.855	[0.708, 1.03]	0.105
Coefficient of Variation	0.990	[0.981, 0.998]	**0.020**
Baseline NLR	0.986	[0.951, 1.02]	0.454
Baseline Haemoglobin Count	0.853	[0.803, 0.906]	**<0.0001**
Baseline Anaemia	1.71	[1.32, 2.21]	**<0.0001**
Hypoglycaemia Frequency	1.04	[0.969, 1.12]	0.263
Anti-Diabetic Agent			
Sulphonylurea	0.924	[0.726, 1.18]	0.522
Biguanide	1.21	[0.948, 1.54]	0.127
DPP4 Inhibitor	0.808	[0.113, 5.76]	0.831
Thiazolidinedione	0.608	[0.271, 1.37]	0.228
Alpha-Glucosidase Inhibitor	1.10	[0.567, 2.14]	0.775
Peripheral Vascular Disease (n=711)			
Age	1.02	[1.01, 1.02]	**<0.0001**
Male	1.36	[1.17, 1.58]	**<0.0001**
Diabetes Duration	0.968	[0.939, 0.998]	**0.035**
HbA1c			
Baseline	1.03	[0.987, 1.07]	0.195

Mean	1.20	[1.14, 1.26]	**<0.0001**
Standard Deviation	1.19	[1.11, 1.28]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.02]	**0.001**
Total Cholesterol			
Baseline	0.938	[0.870, 1.01]	0.095
Mean	1.14	[1.05, 1.25]	**0.003**
Standard Deviation	1.34	[1.21, 1.50]	**<0.0001**
Coefficient of Variation	1.02	[1.01, 1.03]	**<0.0001**
HDL Cholesterol			
Baseline	1.10	[0.905, 1.35]	0.330
Mean	0.452	[0.346, 0.590]	**<0.0001**
Standard Deviation	1.82	[0.840, 3.93]	0.130
Coefficient of Variation	1.03	[1.02, 1.03]	**<0.0001**
LDL Cholesterol			
Baseline	0.937	[0.853, 1.03]	0.177
Mean	1.11	[0.994, 1.24]	0.065
Standard Deviation	1.54	[1.26, 1.88]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.02]	**<0.001**
Triglyceride			
Baseline	1.00	[0.959, 1.05]	0.898
Mean	1.09	[1.06, 1.13]	**<0.0001**
Standard Deviation	1.09	[1.05, 1.13]	**<0.0001**
Coefficient of Variation	1.01	[1.00, 1.01]	**<0.0001**
Baseline NLR	1.02	[1.00, 1.03]	**0.016**
Baseline Haemoglobin Count	0.868	[0.836, 0.901]	**<0.0001**
Baseline Anaemia	1.71	[1.46, 2.01]	**<0.0001**
Hypoglycaemia Frequency	1.07	[1.03, 1.12]	**0.001**
Anti-Diabetic Agent			
Sulphonylurea	0.897	[0.772, 1.04]	0.159
Biguanide	0.768	[0.663, 0.889]	0.889
DPP4 Inhibitor	0.943	[0.304, 2.93]	0.919
Thiazolidinedione	0.952	[0.634, 1.43]	0.812
Alpha-Glucosidase Inhibitor	1.81	[1.30, 2.52]	**<0.001**
Intracranial Haemorrhage (n=1049)			
Age	1.03	[1.02, 1.03]	**<0.0001**
Male	1.20	[1.06, 1.35]	**0.004**
Diabetes Duration	0.966	[0.942, 0.990]	**0.006**
HbA1c			
Baseline	1.02	[0.988, 1.05]	0.218
Mean	0.991	[0.945, 1.04]	0.703
Standard Deviation	1.04	[0.970, 1.12]	0.265
Coefficient of Variation	1.01	[0.998, 1.01]	0.168
Total Cholesterol			
Baseline	0.950	[0.893, 1.01]	0.104

Mean	0.945	[0.875, 1.02]	0.155
Standard Deviation	1.13	[1.00, 1.29]	**0.050**
Coefficient of Variation	1.01	[1.00, 1.02]	**0.026**
HDL Cholesterol			
Baseline	0.980	[0.826, 1.16]	0.815
Mean	0.757	[0.619, 0.926]	**0.007**
Standard Deviation	1.17	[0.598, 2.30]	0.644
Coefficient of Variation	1.01	[1.00, 1.02]	**0.020**
LDL Cholesterol			
Baseline	0.975	[0.902, 1.05]	0.513
Mean	0.875	[0.794, 0.964]	**0.007**
Standard Deviation	1.15	[0.954, 1.38]	0.145
Coefficient of Variation	1.01	[1.00, 1.01]	**0.008**
Triglyceride			
Baseline	0.994	[0.955, 1.04]	0.776
Mean	1.07	[1.03, 1.10]	**<0.001**
Standard Deviation	1.05	[1.00, 1.10]	**0.033**
Coefficient of Variation	1.00	[0.998, 1.01]	0.462
Baseline NLR	1.00	[0.989, 1.02]	0.623
Baseline Haemoglobin Count	0.923	[0.894, 0.952]	**<0.0001**
Baseline Anaemia	1.19	[1.05, 1.35]	**0.008**
Hypoglycaemia Frequency	1.04	[0.997, 1.08]	0.075
Anti-Diabetic Agent			
Sulphonylurea	1.17	[1.04, 1.32]	**0.012**
Biguanide	0.875	[0.775, 0.988]	**0.031**
DPP4 Inhibitor	0.850	[0.318, 2.27]	0.746
Thiazolidinedione	1.05	[0.763, 1.45]	0.764
Alpha-Glucosidase Inhibitor	0.887	[0.610, 1.29]	0.533
Ischemic Stroke and Transient Ischemic Attack/ Transient Ischemic Attack (n=1350)			
Age	1.03	[1.03, 1.04]	**<0.0001**
Male	0.998	[0.897, 1.11]	0.964
Diabetes Duration	0.989	[0.968, 1.01]	0.327
HbA1c			
Baseline	0.984	[0.954, 1.01]	0.291
Mean	1.05	[1.01, 1.09]	**0.024**
Standard Deviation	0.994	[0.929, 1.06]	0.854
Coefficient of Variation	0.997	[0.990, 1.00]	0.350
Total Cholesterol			
Baseline	0.987	[0.935, 1.04]	0.637
Mean	1.03	[0.967, 1.11]	0.329
Standard Deviation	1.10	[0.977, 1.24]	0.114
Coefficient of Variation	1.01	[0.999, 1.01]	0.113
HDL Cholesterol			
Baseline	1.01	[0.872, 1.18]	0.868

Mean	0.800	[0.671, 0.954]	**0.013**
Standard Deviation	1.20	[0.660, 2.18]	0.551
Coefficient of Variation	1.01	[1.00, 1.02]	0.061
LDL Cholesterol			
Baseline	0.967	[0.903, 1.04]	0.335
Mean	1.00	[0.922, 1.09]	0.942
Standard Deviation	1.14	[0.962, 1.34]	0.133
Coefficient of Variation	1.00	[0.999, 1.01]	0.119
Triglyceride			
Baseline	0.981	[0.943, 1.02]	0.342
Mean	1.05	[1.02, 1.09]	**0.003**
Standard Deviation	1.03	[0.985, 1.08]	0.193
Coefficient of Variation	1.00	[0.997, 1.00]	0.995
Baseline NLR	1.01	[0.992, 1.02]	0.404
Baseline Haemoglobin Count	0.986	[0.958, 1.02]	0.337
Baseline Anaemia	1.07	[0.956, 1.20]	0.232
Hypoglycaemia Frequency	1.00	[0.963, 1.04]	0.963
Anti-Diabetic Agent			
Sulphonylurea	1.19	[1.07, 1.33]	**0.001**
Biguanide	1.18	[1.06, 1.32]	**0.003**
DPP4 Inhibitor	0.491	[0.158, 1.52]	0.218
Thiazolidinedione	0.765	[0.551, 1.06]	0.107
Alpha-Glucosidase Inhibitor	0.891	[0.640, 1.24]	0.496
Ischemic Heart Disease (n=4532)			
Age	1.02	[1.02, 1.02]	**<0.0001**
Male	1.03	[0.975, 1.10]	0.266
Diabetes Duration	1.02	[1.01, 1.04]	**<0.0001**
HbA1c			
Baseline	0.996	[0.980, 1.01]	0.647
Mean	1.05	[1.03, 1.08]	**<0.0001**
Standard Deviation	0.987	[0.951, 1.02]	0.470
Coefficient of Variation	0.996	[0.993, 1.00]	**0.046**
Total Cholesterol			
Baseline	1.01	[0.978, 1.04]	0.634
Mean	1.25	[1.21, 1.30]	**<0.0001**
Standard Deviation	1.22	[1.16, 1.29]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.01]	**<0.0001**
HDL Cholesterol			
Baseline	1.12	[1.03, 1.21]	**0.006**
Mean	0.642	[0.581, 0.710]	**<0.0001**
Standard Deviation	0.596	[0.421, 0.844]	**0.004**
Coefficient of Variation	1.00	[0.998, 1.01]	0.366
LDL Cholesterol			
Baseline	1.01	[0.972, 1.05]	0.675

Mean	1.27	[1.22, 1.33]	**<0.0001**
Standard Deviation	1.43	[1.32, 1.56]	**<0.0001**
Coefficient of Variation	1.01	[1.00, 1.01]	**<0.0001**
Triglyceride			
Baseline	0.981	[0.960, 1.00]	0.084
Mean	1.09	[1.07, 1.10]	**<0.0001**
Standard Deviation	1.05	[1.02, 1.07]	**<0.0001**
Coefficient of Variation	1.00	[0.999, 1.00]	0.313
Baseline NLR	1.00	[0.996, 1.01]	0.379
Baseline Haemoglobin Count	0.925	[0.911, 0.939]	**<0.0001**
Baseline Anaemia	1.29	[1.21, 1.38]	**<0.0001**
Hypoglycaemia Frequency	1.00	[0.979, 1.02]	0.983
Anti-Diabetic Agent			
Sulphonylurea	1.16	[1.09, 1.23]	**<0.0001**
Biguanide	1.10	[1.03, 1.16]	**0.003**
DPP4 Inhibitor	1.13	[0.752, 1.71]	0.553
Thiazolidinedione	1.13	[0.974, 1.31]	0.107
Alpha-Glucosidase Inhibitor	1.20	[1.03, 1.41]	**0.023**
Acute Myocardial Infarction (n=3178)			
Age	1.03	[1.03, 1.03]	**<0.0001**
Male	0.80	[0.914, 1.05]	0.562
Diabetes Duration	0.989	[0.975, 1.00]	0.103
HbA1c			
Baseline	1.01	[0.986, 1.03]	0.603
Mean	1.07	[1.04, 1.10]	**<0.0001**
Standard Deviation	1.01	[0.966, 1.05]	0.726
Coefficient of Variation	0.998	[0.994, 1.00]	0.338
Total Cholesterol			
Baseline	0.990	[0.956, 1.03]	0.562
Mean	1.18	[1.13, 1.23]	**<0.0001**
Standard Deviation	1.22	[1.14, 1.30]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.02]	**<0.0001**
HDL Cholesterol			
Baseline	0.992	[0.900, 1.09]	0.867
Mean	0.522	[0.462, 0.591]	**<0.0001**
Standard Deviation	0.680	[0.453, 1.02]	0.064
Coefficient of Variation	1.01	[1.00, 1.01]	**0.005**
LDL Cholesterol			
Baseline	0.982	[0.940, 1.03]	0.430
Mean	1.18	[1.12, 1.24]	**<0.0001**
Standard Deviation	1.41	[1.28, 1.56]	**<0.0001**
Coefficient of Variation	1.01	[1.00, 1.01]	**<0.0001**
Triglyceride			
Baseline	0.977	[0.952, 1.00]	0.076

Mean	1.09	[1.08, 1.11]	**<0.0001**
Standard Deviation	1.06	[1.03, 1.08]	**<0.0001**
Coefficient of Variation	1.00	[1.00, 1.00]	**0.042**
Baseline NLR	1.00	[0.993, 1.01]	0.601
Baseline Haemoglobin Count	0.895	[0.879, 0.911]	**<0.0001**
Baseline Anaemia	1.48	[1.38, 1.60]	**<0.0001**
Hypoglycaemia Frequency	1.01	[0.982, 1.03]	0.619
Anti-Diabetic Agent			
Sulphonylurea	1.22	[1.14, 1.31]	**<0.0001**
Biguanide	0.961	[0.896, 1.03]	0.271
DPP4 Inhibitor	0.836	[0.474, 1.47]	0.535
Thiazolidinedione	0.868	[0.710, 1.06]	0.164
Alpha-Glucosidase Inhibitor	1.23	[1.02, 1.48]	**0.032**
Atrial Fibrillation (n=1846)			
Age	1.04	[1.04, 1.05]	**<0.0001**
Male	0.884	[0.807, 0.968]	**0.008**
Diabetes Duration	0.989	[0.975, 1.00]	0.103
HbA1c			
Baseline	1.00	[0.975, 1.03]	0.975
Mean	0.961	[0.926, 0.996]	**0.029**
Standard Deviation	1.05	[0.993, 1.11]	0.090
Coefficient of Variation	1.01	[1.00, 1.01]	**0.024**
Total Cholesterol			
Baseline	1.03	[0.990, 1.08]	0.134
Mean	0.889	[0.838, 0.943]	**<0.0001**
Standard Deviation	1.12	[1.02, 1.23]	**0.020**
Coefficient of Variation	1.01	[1.01, 1.02]	**<0.001**
HDL Cholesterol			
Baseline	1.02	[0.894, 1.15]	0.823
Mean	0.667	[0.572, 0.779]	**<0.0001**
Standard Deviation	0.988	[0.590, 1.65]	0.962
Coefficient of Variation	1.01	[1.00, 1.02]	**0.005**
LDL Cholesterol			
Baseline	1.03	[0.969, 1.09]	0.389
Mean	0.832	[0.773, 0.895]	**<0.0001**
Standard Deviation	1.20	[1.04, 1.37]	**0.011**
Coefficient of Variation	1.01	[1.01, 1.01]	**<0.0001**
Triglyceride			
Baseline	0.982	[0.950, 1.01]	0.266
Mean	1.04	[1.01, 1.07]	**0.019**
Standard Deviation	0.998	[0.955, 1.04]	0.914
Coefficient of Variation	0.998	[0.995, 1.00]	0.147
Baseline NLR	1.00	[0.992, 1.02]	0.574

Baseline Haemoglobin Count	0.907	[0.886, 0.929]	**<0.0001**
Baseline Anaemia	1.41	[1.28, 1.56]	**<0.0001**
Hypoglycaemia Frequency	1.03	[1.00, 1.06]	**0.030**
Anti-Diabetic Agent			
Sulphonylurea	1.17	[1.07, 1.28]	**0.001**
Biguanide	0.933	[0.851, 1.02]	0.136
DPP4 Inhibitor	0.716	[0.321, 1.60]	0.413
Thiazolidinedione	0.965	[0.751, 1.24]	0.781
Alpha-Glucosidase Inhibitor	1.00	[0.765, 1.31]	0.997
Heart Failure (n=1810)			
Age	1.03	[1.03, 1.04]	**<0.0001**
Male	1.27	[1.17, 1.39]	**<0.0001**
Diabetes Duration	0.937	[0.919, 0.955]	**<0.0001**
HbA1c			
Baseline	1.02	[0.993, 1.04]	0.170
Mean	1.05	[1.02, 1.09]	**0.004**
Standard Deviation	1.10	[1.04, 1.16]	**<0.001**
Coefficient of Variation	1.01	[1.00, 1.01]	**0.001**
Total Cholesterol			
Baseline	0.999	[0.954, 1.05]	0.961
Mean	1.06	[0.997, 1.12]	0.065
Standard Deviation	1.28	[1.19, 1.38]	**<0.0001**
Coefficient of Variation	1.02	[1.01, 1.03]	**<0.0001**
HDL Cholesterol			
Baseline	0.976	[0.857, 1.11]	0.709
Mean	0.417	[0.352, 0.494]	**<0.0001**
Standard Deviation	1.25	[0.754, 2.07]	0.387
Coefficient of Variation	1.02	[1.02, 1.03]	**<0.0001**
LDL Cholesterol			
Baseline	1.02	[0.962, 1.08]	0.517
Mean	1.06	[0.988, 1.14]	0.103
Standard Deviation	1.51	[1.33, 1.71]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.02]	**<0.0001**
Triglyceride			
Baseline	1.00	[0.973, 1.03]	0.930
Mean	1.09	[1.07, 1.12]	**<0.0001**
Standard Deviation	1.06	[1.03, 1.09]	**<0.001**
Coefficient of Variation	1.00	[1.00, 1.01]	**0.006**
Baseline NLR	1.02	[1.01, 1.03]	**<0.0001**
Baseline Haemoglobin Count	0.835	[0.815, 0.854]	**<0.0001**
Baseline Anaemia	1.85	[1.67, 2.04]	**<0.0001**
Hypoglycaemia Frequency	1.08	[1.06, 1.11]	**<0.0001**

Anti-Diabetic Agent			
Sulphonylurea	1.30	[1.19, 1.43]	**<0.0001**
Biguanide	0.618	[0.563, 0.678]	**<0.0001**
DPP4 Inhibitor	0.608	[0.253, 1.46]	0.266
Thiazolidinedione	0.722	[0.540, 0.964]	**0.027**
Alpha-Glucosidase Inhibitor	1.17	[0.905, 1.50]	0.235
All-Cause Mortality			
Age	1.06	[1.06, 1.06]	**<0.0001**
Male	1.14	[1.10, 1.18]	**<0.0001**
Diabetes Duration	0.930	[0.923, 0.936]	**<0.0001**
HbA1c			
Baseline	0.997	[0.988, 1.01]	0.549
Mean	0.964	[0.951, 0.977]	**<0.0001**
Standard Deviation	1.11	[1.09, 1.14]	**<0.0001**
Coefficient of Variation	1.01	[1.01, 1.01]	**<0.0001**
Total Cholesterol			
Baseline	0.992	[0.976, 1.01]	0.340
Mean	0.965	[0.943, 0.988]	**0.003**
Standard Deviation	1.29	[1.25, 1.33]	**<0.0001**
Coefficient of Variation	1.02	[1.02, 1.03]	**<0.0001**
HDL Cholesterol			
Baseline	0.964	[0.921, 1.01]	0.114
Mean	0.571	[0.537, 0.607]	**<0.0001**
Standard Deviation	3.11	[2.60, 3.72]	**<0.0001**
Coefficient of Variation	0.970	[0.927, 1.02]	0.194
LDL Cholesterol			
Baseline	1.02	[0.995, 1.04]	0.131
Mean	0.919	[0.893, 0.946]	**<0.0001**
Standard Deviation	1.56	[1.48, 1.64]	**<0.0001**
Coefficient of Variation	1.02	[1.01, 1.02]	**<0.0001**
Triglyceride			
Baseline	0.989	[0.978, 1.00]	0.0557
Mean	1.06	[1.05, 1.07]	**<0.0001**
Standard Deviation	1.03	[1.02, 1.04]	**<0.0001**
Coefficient of Variation	1.00	[1.00, 1.00]	**0.002**
Baseline NLR	1.03	[1.03, 1.04]	**<0.0001**
Baseline Haemoglobin Count	0.794	[0.786, 0.801]	**<0.0001**
Baseline Anaemia	2.11	[2.04, 2.20]	**<0.0001**
Hypoglycaemia Frequency	1.08	[1.06, 1.09]	**<0.0001**
Anti-Diabetic Agent			
Sulphonylurea	1.13	[1.09, 1.17]	**<0.0001**
Biguanide	0.541	[0.522, 0.560]	**<0.0001**
DPP4 Inhibitor	0.450	[0.324, 0.655]	**<0.0001**
Thiazolidinedione	0.703	[0.632, 0.782]	**<0.0001**

Alpha-Glucosidase Inhibitor	1.05	[0.953, 1.17]	0.305

3.3.4. The relationship between neutrophil-lymphocyte ratio, frequency of hypoglycaemic episodes and glycaemic variability

The average number of hypoglycaemic episodes experienced is 0.537 ± 1.38, and the mean baseline NLR is 3.80 ± 4.16. The baseline mean value of HbA1c was 8.56 ± 1.94%. Variability, represented by SD and CV, are 1.28 ± 0.851 and 14.5 ± 8.76 respectively. HbA1c and lipid variability were significantly associated with baseline NLR with cancer status and aspirin use adjusted, and the associations were summarized in **Table 10.** Similarly, HbA1c variability was also found to be positively correlated with hypoglycaemic frequency (SD: OR= 1.13, 95% CI= [1.12, 1.16], $p < 0.0001$; CV: OR= 1.02, 95% CI= [1.02, 1.02], $p < 0.0001$). Additionally, triglyceride SD is positively correlated with both LDL-C (SD: OR= 1.86, 95% CI= [1.78, 1.93], $p < 0.0001$; CV: OR= 1.02, 95% CI= [1.02, 1.02], $p < 0.0001$) and HDL-C (OR= 2.92, 95% CI= [2.48, 3.43], $p < 0.0001$) variability. After the exclusion of calculated LDL-C measurements, the significant association between LDL-C variability and triglyceride SD remains (SD: OR= 1.90, 95% CI= [1.79, 2.02], $p < 0.0001$; CV: OR= 1.02, 95% CI= [1.02, 1.02], $p < 0.0001$).

Table 10. Significant associations between HbA1c/ lipid variability with baseline neutrophil-lymphocyte ratio

HbA1c/ Lipid Variability	Hazard ratio [95% Confidence Interval]	P-Value
HbA1c: SD	1.01 [1.01, 1.01]	< 0.0001
HbA1c: CV	1.13 [1.10, 1.17]	< 0.0001
HDL-C: SD	1.00 [1.00, 1.00]	< 0.0001
HDL-C: CV	1.19 [1.15, 1.23]	< 0.0001
Triglyceride: CV	1.08 [1.01, 1.16]	0.019
Total Cholesterol: SD	1.01 [1.00, 1.01]	< 0.0001
Total Cholesterol: CV	1.10 [1.07, 1.13]	< 0.0001

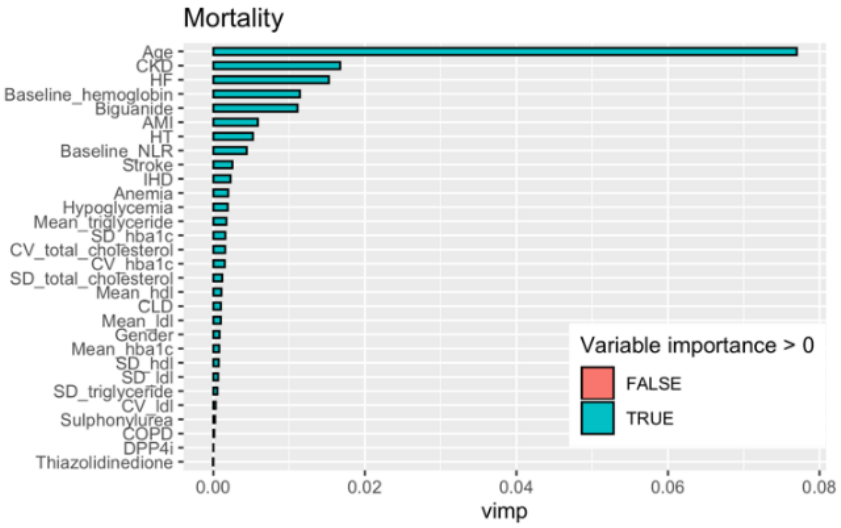
SD: standard deviation; CV: coefficient of variation; HDL-C: high density lipoprotein-cholesterol;

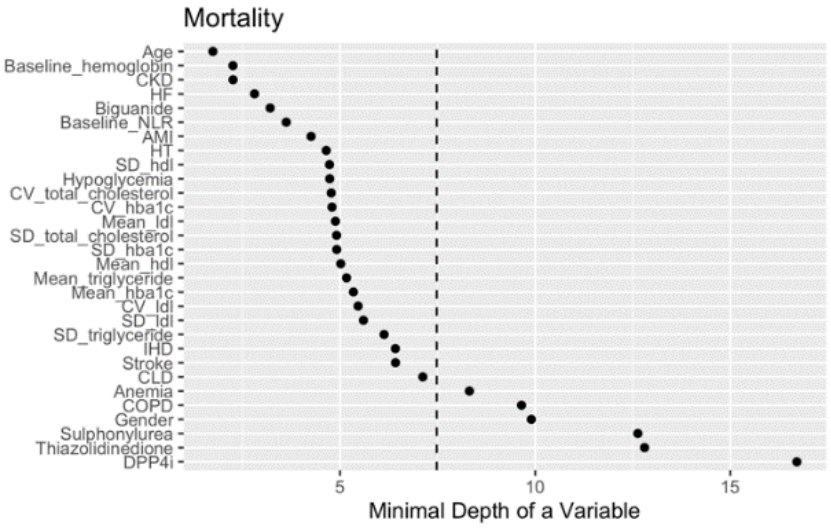
The analysis was adjusted to cancer status and aspirin use.

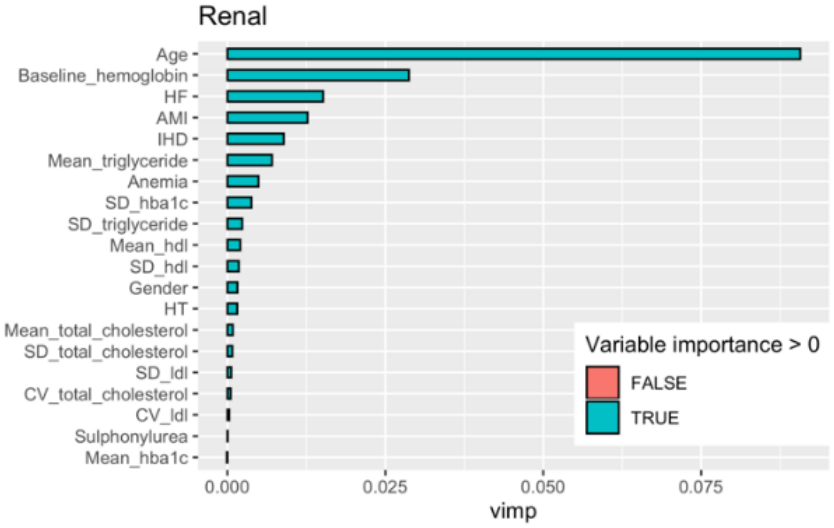
3.3.5. Survival learning results

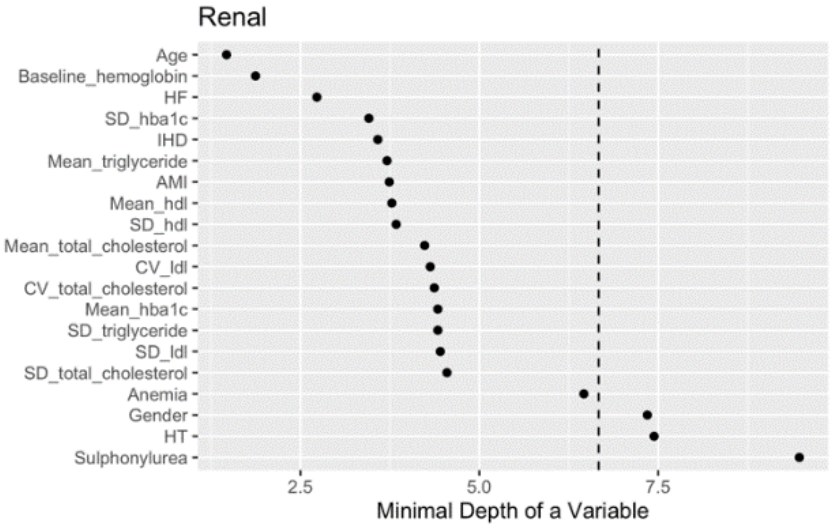
A regularized and weighted RSF model was devised, with significant variables identified from univariate Cox regression inputted. This yielded the importance ranking and minimal depth of each variable in the tree structure of the model, as shown in **Figure 4**. The corresponding decision rules derived by using the regularized and weighted RSF model were generated based on the out-of-bag (OOB) validation dataset (n=5 037; **Figure 5**). The minimal depth assumes that variables with a high impact on the prediction are those that most frequently split nodes nearest to the root node, where they partition the largest samples of the population. Minimal depth measures important risk factors by averaging the depth of the first split for each variable over all trees within the forest. Smaller minimal depth values indicate that the variable separates large groups of observations, and therefore has a large impact on the prediction. Both importance ranking and minimal depth were used to uncover the most important variables for predicting time-to-event complication outcomes.

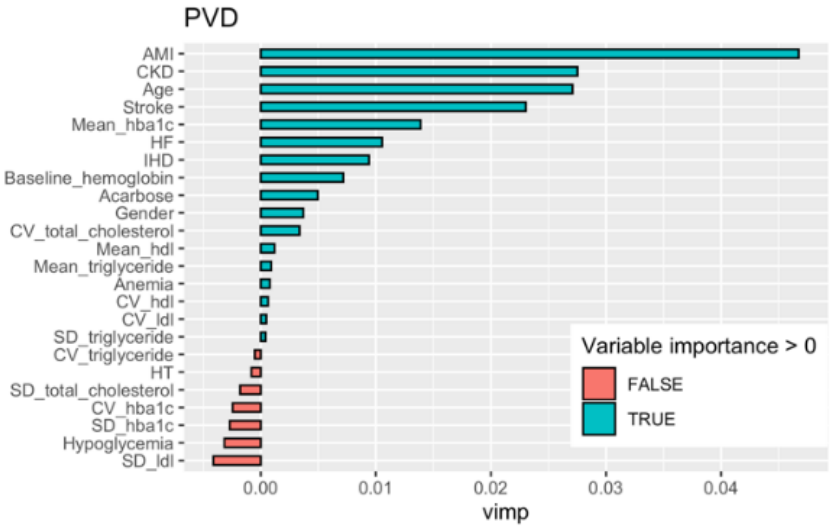
Figure 4. Importance ranking and minimal depth of significant univariable variables to predict mortality and complications using regularized and weighted random survival forest model

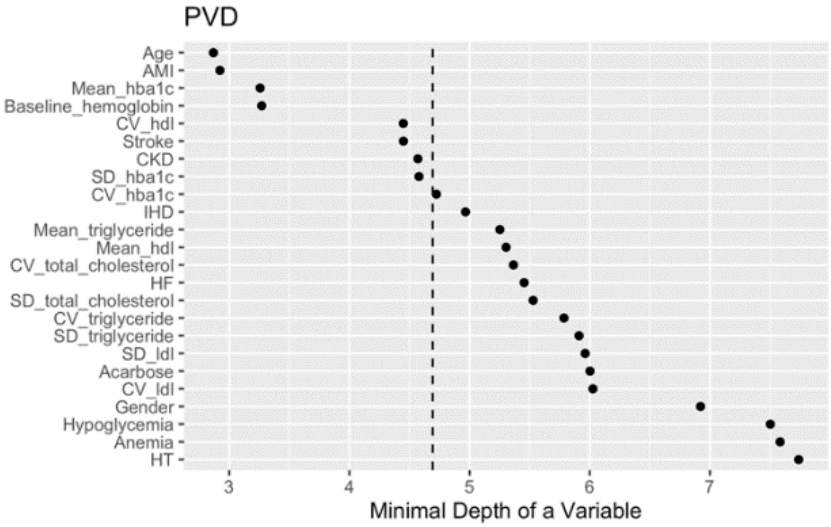


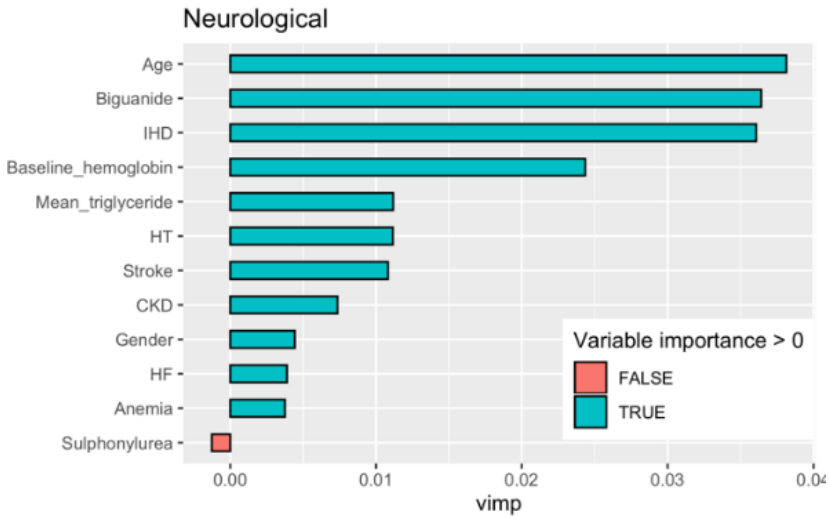


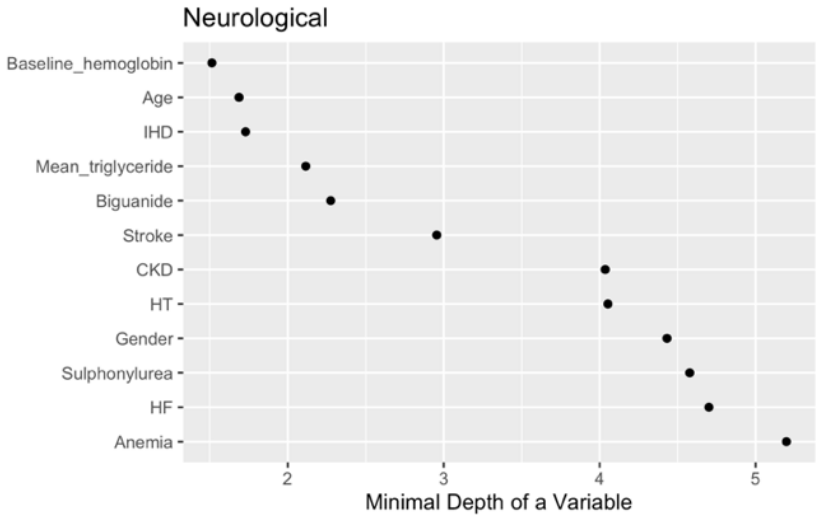


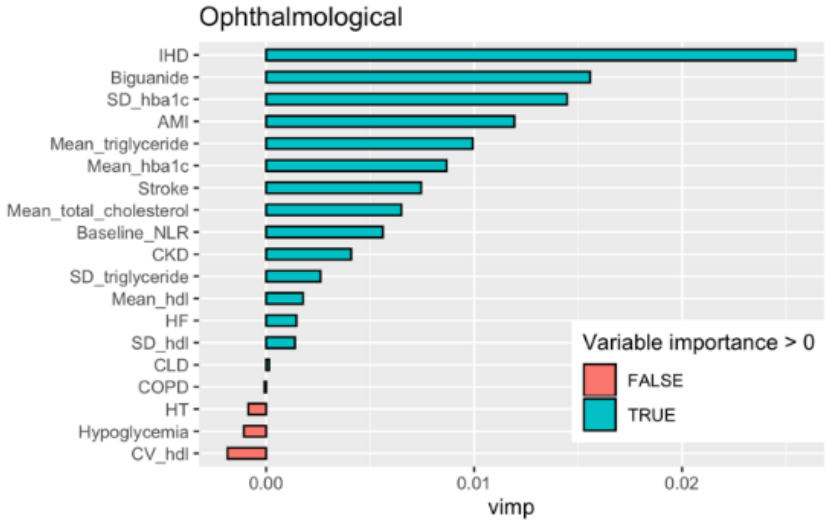

Renal
Age
Baseline_hemoglobin
HF
SD_hba1c
IHD
Mean_triglyceride
AMI
Mean_hdl
SD_hdl
Mean_total_cholesterol
CV_ldl
CV_total_cholesterol
Mean_hba1c
SD_triglyceride
SD_ldl
SD_total_cholesterol
Anemia
Gender
HT
Sulphonylurea
2.5
5.0
7.5
Minimal Depth of a Variable

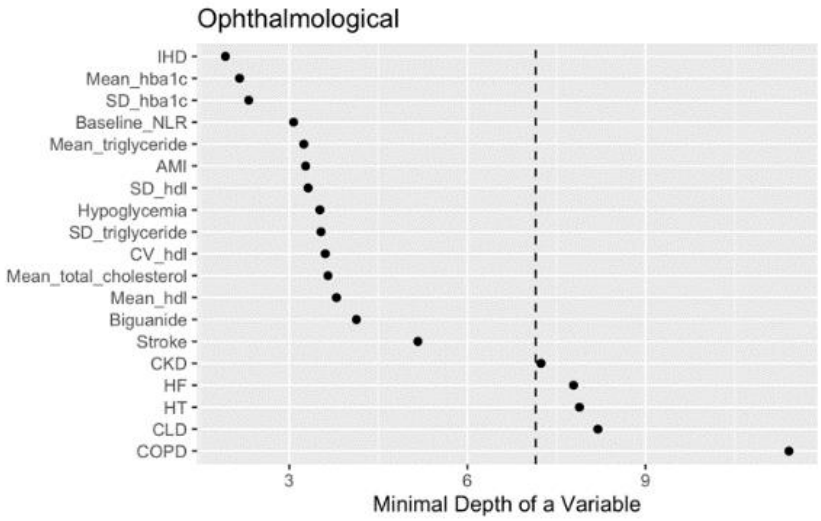

PVD
AMI
CKD
Age
Stroke
Mean_hba1c
HF
IHD
Baseline_hemoglobin
Acarbose
Gender
CV_total_cholesterol
Mean_hdl
Mean_triglyceride
Anemia
CV_hdl
CV_ldl
SD_triglyceride
CV_triglyceride
HT
SD_total_cholesterol
CV_hba1c
SD_hba1c
Hypoglycemia
SD_ldl
Variable importance > 0
FALSE
TRUE
0.00
0.01
0.02
0.03
0.04
vimp

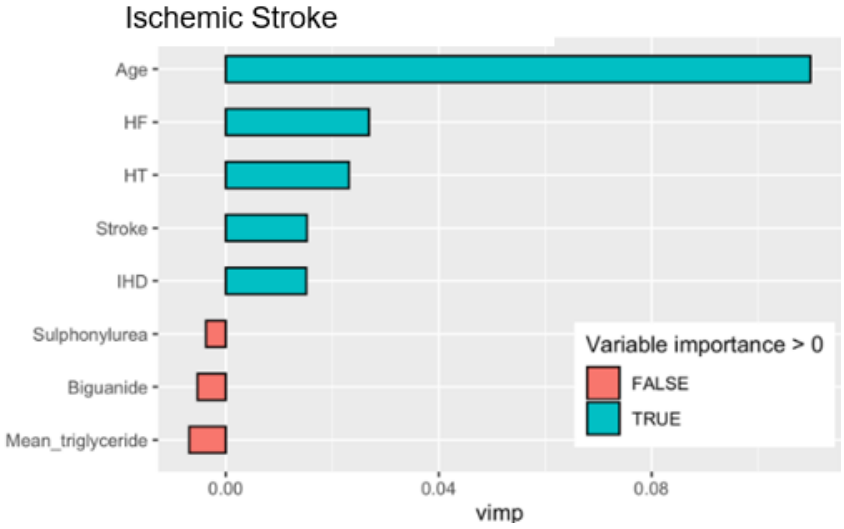

PVD
Age
AMI
Mean_hba1c
Baseline_hemoglobin
CV_hdl
Stroke
CKD
SD_hba1c
CV_hba1c
IHD
Mean_triglyceride
Mean_hdl
CV_total_cholesterol
HF
SD_total_cholesterol
CV_triglyceride
SD_triglyceride
SD_ldl
Acarbose
CV_ldl
Gender
Hypoglycemia
Anemia
HT
3
4
5
6
7
Minimal Depth of a Variable

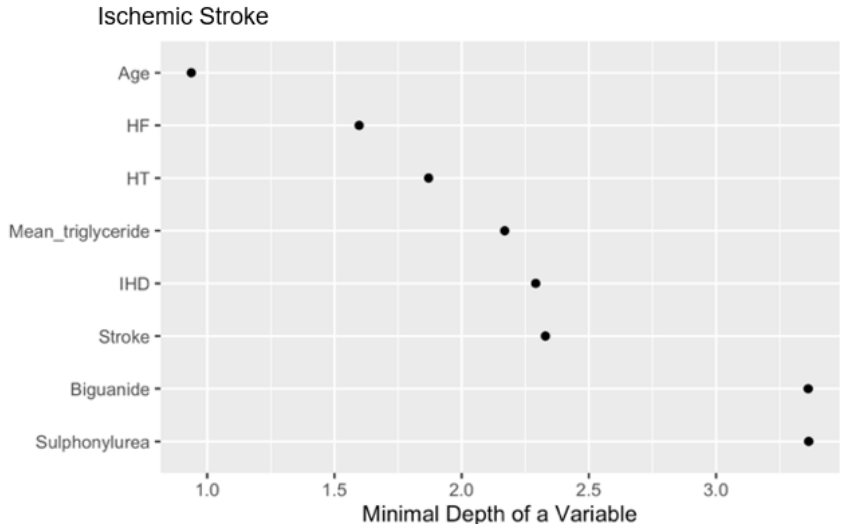

Neurological
Age
Biguanide
IHD
Baseline_hemoglobin
Mean_triglyceride
HT
Stroke
CKD
Gender
HF
Anemia
Sulphonylurea
Variable importance > 0
FALSE
TRUE
0.00
0.01
0.02
0.03
vimp

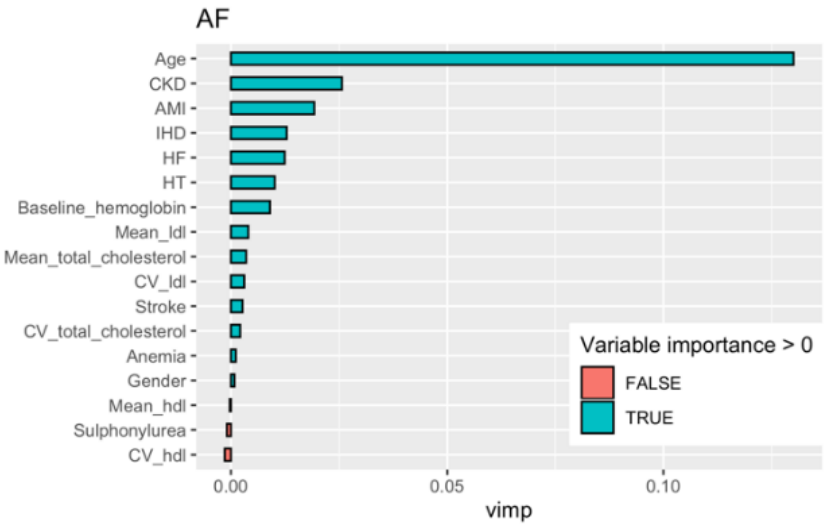

Neurological
Baseline_hemoglobin
Age
IHD
Mean_triglyceride
Biguanide
Stroke
CKD
HT
Gender
Sulphonylurea
HF
Anemia
2
3
4
5
Minimal Depth of a Variable

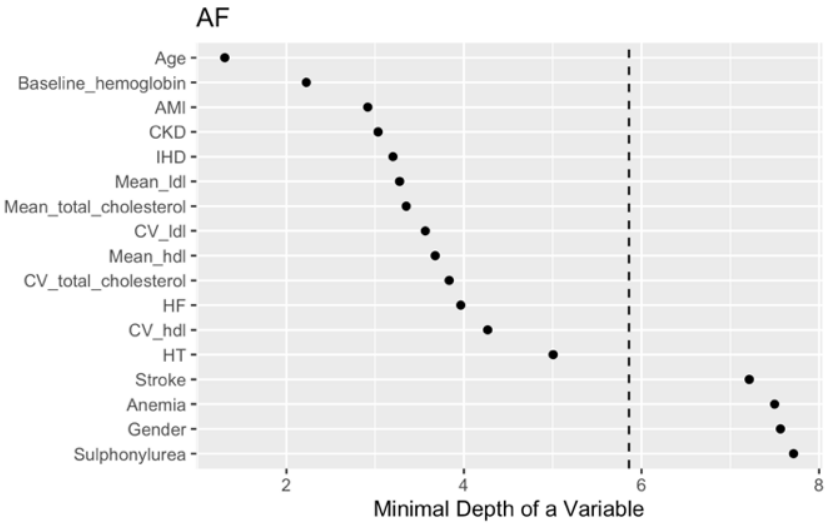

Ophthalmological
IHD
Biguanide
SD_hba1c
AMI
Mean_triglyceride
Mean_hba1c
Stroke
Mean_total_cholesterol
Baseline_NLR
CKD
SD_triglyceride
Mean_hdl
HF
SD_hdl
CLD
COPD
HT
Hypoglycemia
CV_hdl
Variable importance > 0
FALSE
TRUE
0.00
0.01
0.02
vimp

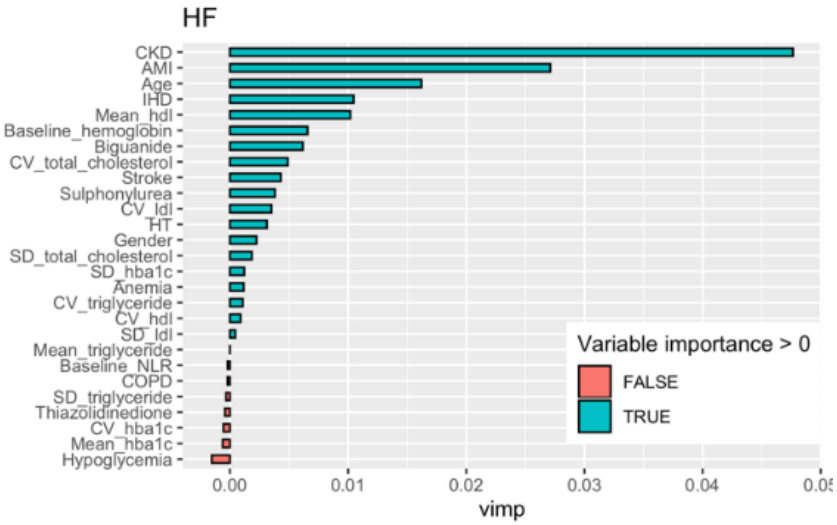

Ophthalmological
IHD
Mean_hba1c
SD_hba1c
Baseline_NLR
Mean_triglyceride
AMI
SD_hdl
Hypoglycemia
SD_triglyceride
CV_hdl
Mean_total_cholesterol
Mean_hdl
Biguanide
Stroke
CKD
HF
HT
CLD
COPD
3
6
9
Minimal Depth of a Variable

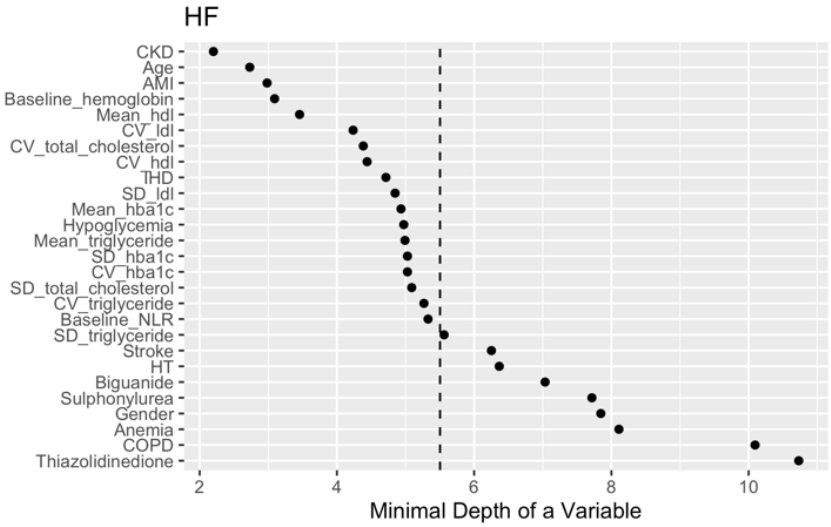

Ischemic Stroke
Age
HF
HT
Stroke
IHD
Sulphonylurea
Biguanide
Mean_triglyceride
Variable importance > 0
FALSE
TRUE
0.00
0.04
0.08
vimp

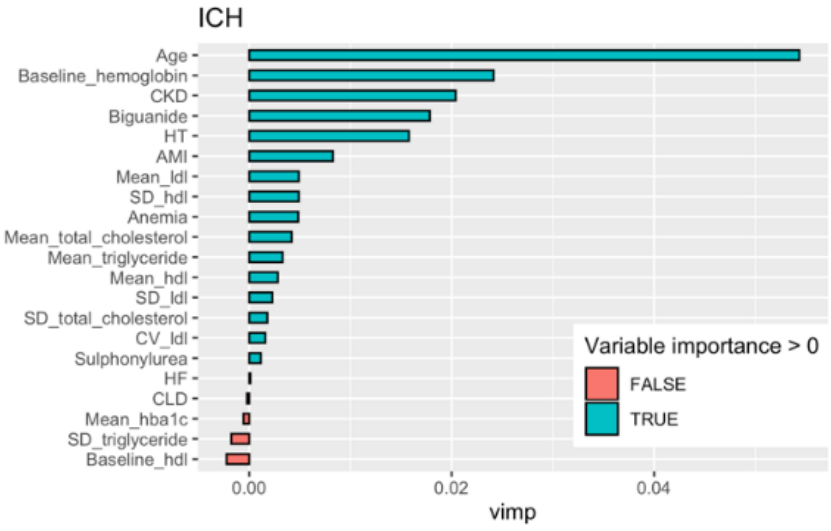

Ischemic Stroke
Age
HF
HT
Mean_triglyceride
IHD
Stroke
Biguanide
Sulphonylurea
1.0
1.5
2.0
2.5
3.0
Minimal Depth of a Variable

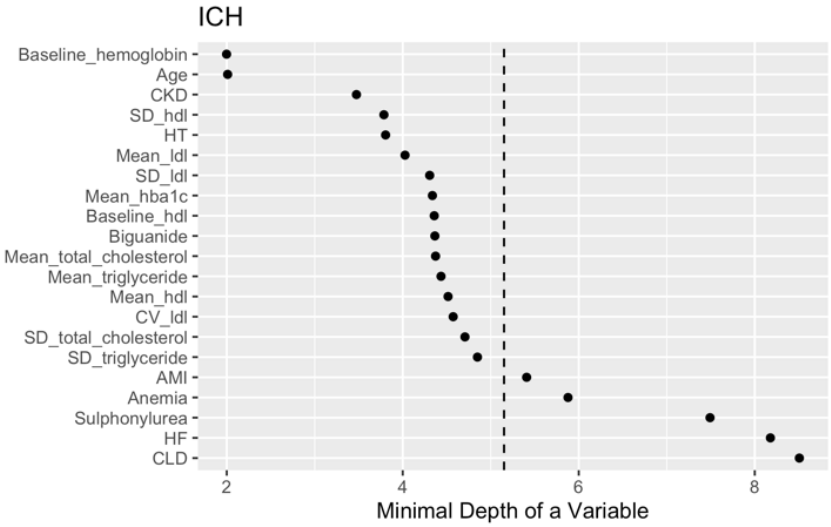

AF
Age
CKD
AMI
IHD
HF
HT
Baseline_hemoglobin
Mean_ldl
Mean_total_cholesterol
CV_ldl
Stroke
CV_total_cholesterol
Anemia
Gender
Mean_hdl
Sulphonylurea
CV_hdl
Variable importance > 0
FALSE
TRUE
0.00
0.05
0.10
vimp

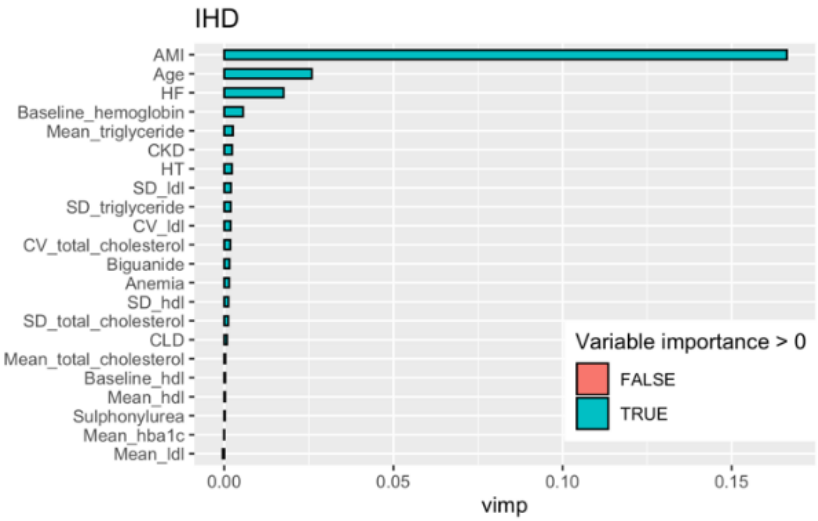

AF
Age
Baseline_hemoglobin
AMI
CKD
IHD
Mean_ldl
Mean_total_cholesterol
CV_ldl
Mean_hdl
CV_total_cholesterol
HF
CV_hdl
HT
Stroke
Anemia
Gender
Sulphonylurea
2
4
6
8
Minimal Depth of a Variable

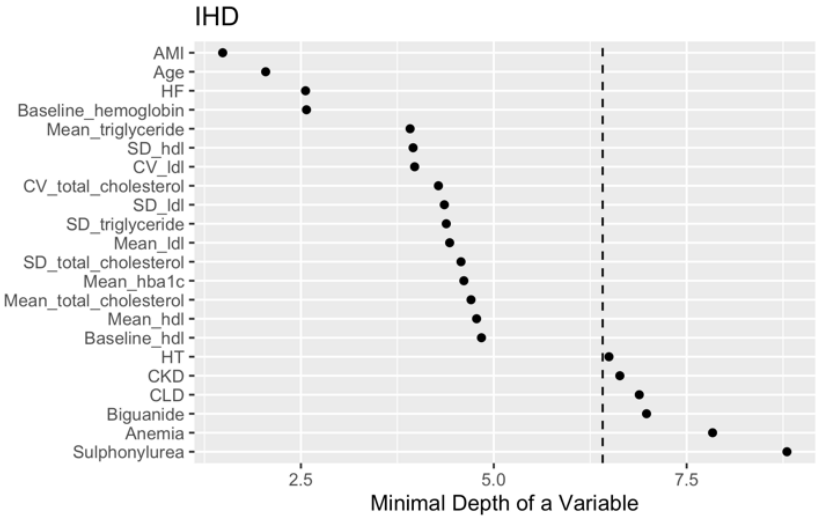

HF
CKD
AMI
Age
IHD
Mean_hdl
Baseline_hemoglobin
Biguanide
CV_total_cholesterol
Stroke
Sulphonylurea
CV_ldl
HT
Gender
SD_total_cholesterol
SD_hba1c
Anemia
CV_triglyceride
CV_hdl
SD_ldl
Mean_triglyceride
Baseline_NLR
COPD
SD_triglyceride
Thiazolidinedione
CV_hba1c
Mean_hba1c
Hypoglycemia
Variable importance > 0
FALSE
TRUE
0.00
0.01
0.02
0.03
0.04
vimp

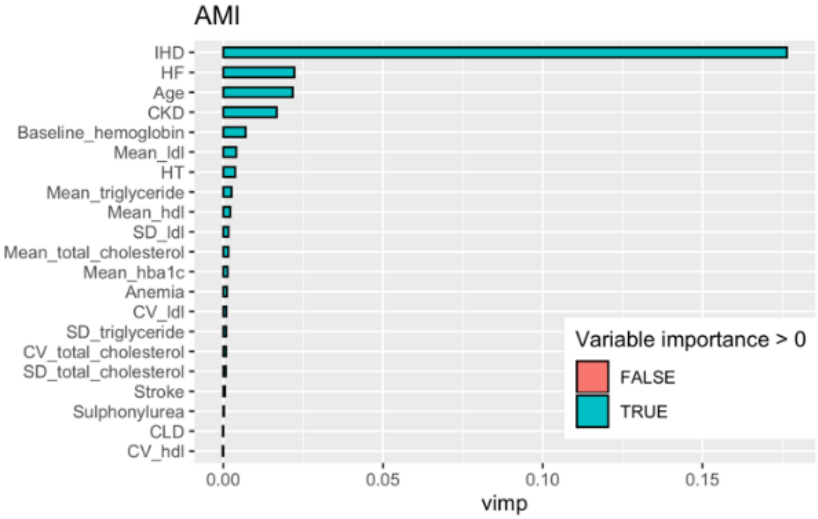

HF
CKD
Age
AMI
Baseline_hemoglobin
Mean_hdl
CV_ldl
CV_total_cholesterol
CV_hdl
THD
SD_ldl
Mean_hba1c
Hypoglycemia
Mean_triglyceride
SD_hba1c
CV_hba1c
SD_total_cholesterol
CV_triglyceride
Baseline_NLR
SD_triglyceride
Stroke
HT
Biguanide
Sulphonylurea
Gender
Anemia
COPD
Thiazolidinedione
2
4
6
8
10
Minimal Depth of a Variable

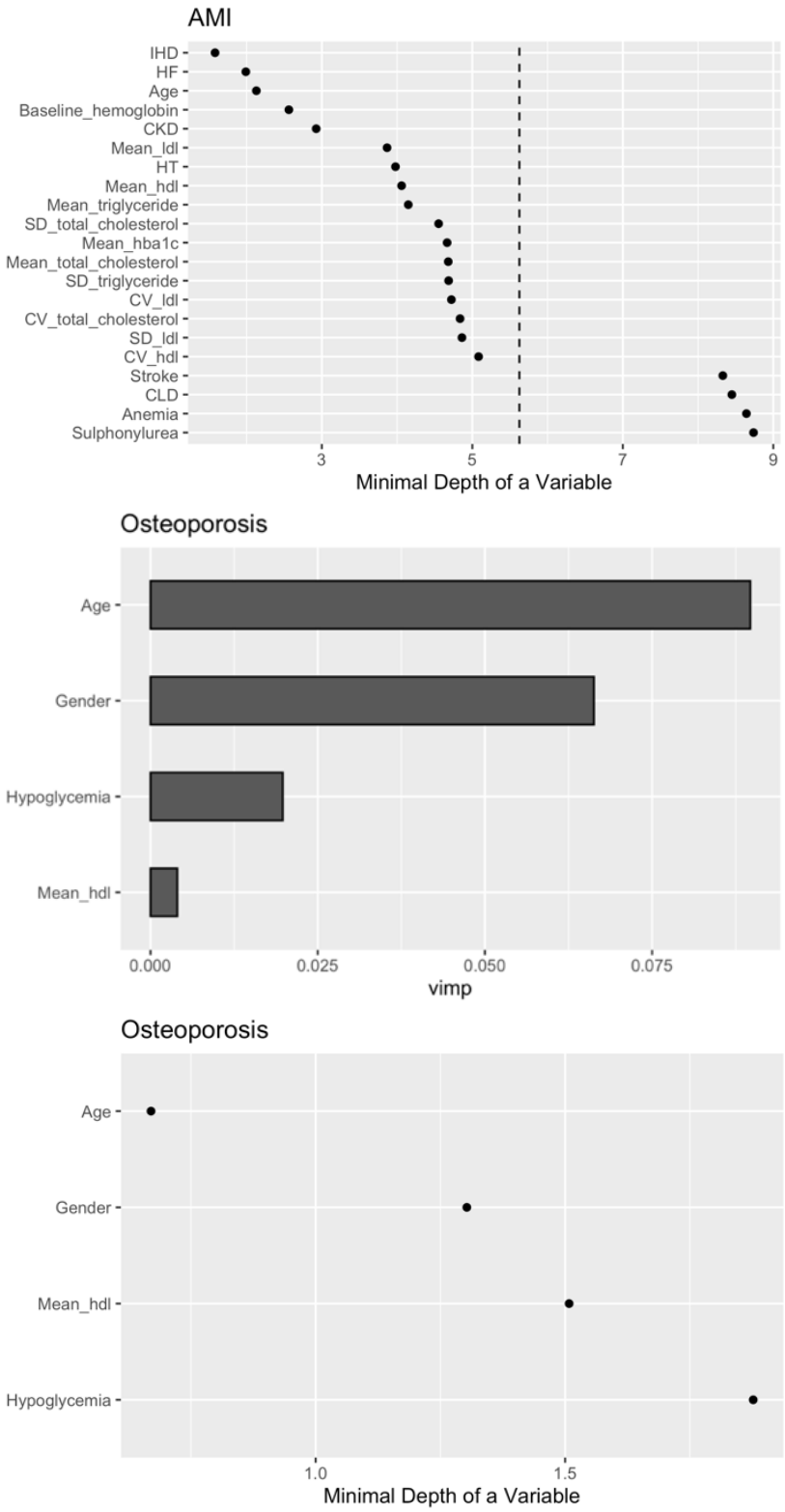

ICH
Age
Baseline_hemoglobin
CKD
Biguanide
HT
AMI
Mean_ldl
SD_hdl
Anemia
Mean_total_cholesterol
Mean_triglyceride
Mean_hdl
SD_ldl
SD_total_cholesterol
CV_ldl
Sulphonylurea
HF
CLD
Mean_hba1c
SD_triglyceride
Baseline_hdl
Variable importance > 0
FALSE
TRUE
0.00
0.02
0.04
vimp


ICH
Baseline_hemoglobin
Age
CKD
SD_hdl
HT
Mean_ldl
SD_ldl
Mean_hba1c
Baseline_hdl
Biguanide
Mean_total_cholesterol
Mean_triglyceride
Mean_hdl
CV_ldl
SD_total_cholesterol
SD_triglyceride
AMI
Anemia
Sulphonylurea
HF
CLD
2
4
6
8
Minimal Depth of a Variable


IHD
AMI
Age
HF
Baseline_hemoglobin
Mean_triglyceride
CKD
HT
SD_ldl
SD_triglyceride
CV_ldl
CV_total_cholesterol
Biguanide
Anemia
SD_hdl
SD_total_cholesterol
CLD
Mean_total_cholesterol
Baseline_hdl
Mean_hdl
Sulphonylurea
Mean_hba1c
Mean_ldl
Variable importance > 0
FALSE
TRUE
0.00
0.05
0.10
0.15
vimp

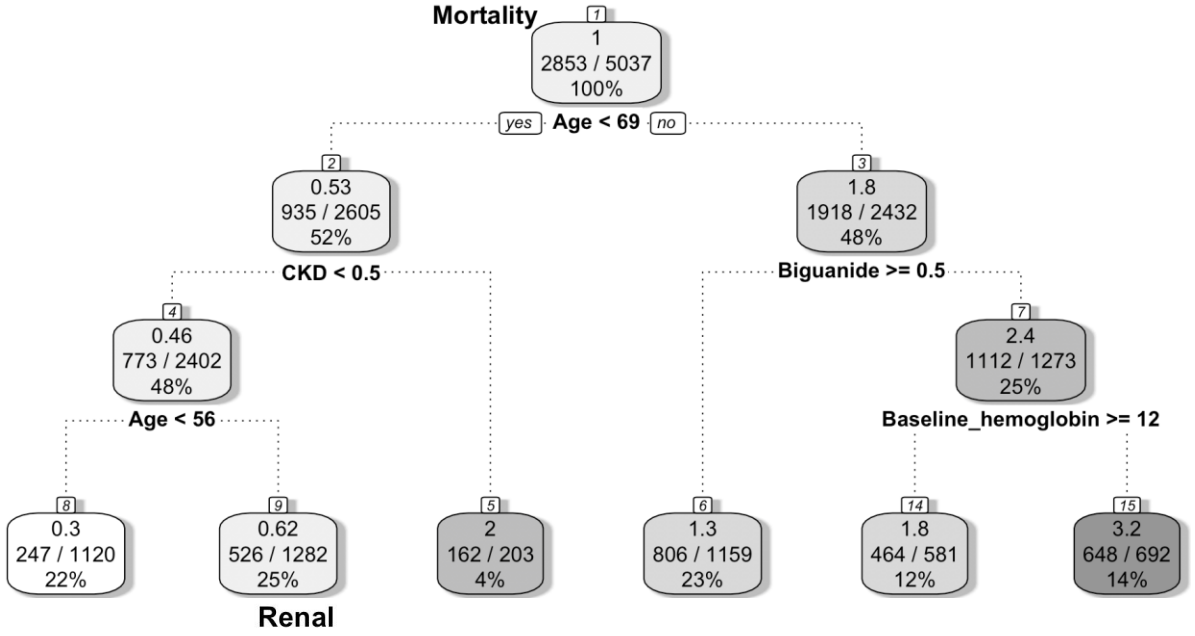

IHD
AMI
Age
HF
Baseline_hemoglobin
Mean_triglyceride
SD_hdl
CV_ldl
CV_total_cholesterol
SD_ldl
SD_triglyceride
Mean_ldl
SD_total_cholesterol
Mean_hba1c
Mean_total_cholesterol
Mean_hdl
Baseline_hdl
HT
CKD
CLD
Biguanide
Anemia
Sulphonylurea
2.5
5.0
7.5
Minimal Depth of a Variable


AMI
IHD
HF
Age
CKD
Baseline_hemoglobin
Mean_ldl
HT
Mean_triglyceride
Mean_hdl
SD_ldl
Mean_total_cholesterol
Mean_hba1c
Anemia
CV_ldl
SD_triglyceride
CV_total_cholesterol
SD_total_cholesterol
Stroke
Sulphonylurea
CLD
CV_hdl
Variable importance > 0
FALSE
TRUE
0.00
0.05
0.10
0.15
vimp

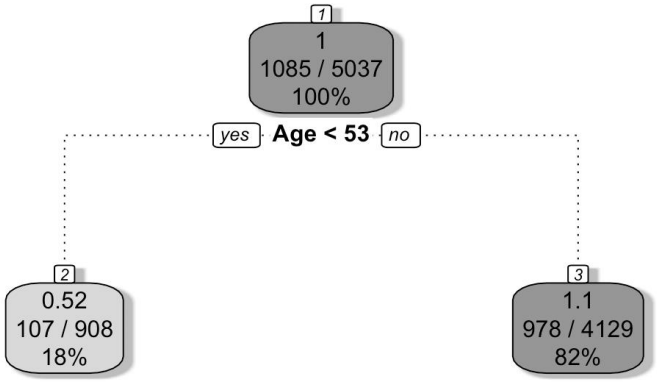


CKD: chronic kidney disease; HF: heart failure; AMI: acute myocardial infarction; HT: hypertension; IHD: ischaemic heart disease; SD: standard deviation; CV: coefficient of variation; HDL: high density lipoprotein cholesterol; LDL: low density lipoprotein cholesterol; COPD: chronic obstructive pulmonary disease; DPP4Ii: dipeptidyl peptidase-4 inhibitor; PVD: peripheral vascular disease; AF: atrial fibrillation; ICH: intracranial haemorrhage; CLD: chronic liver disease

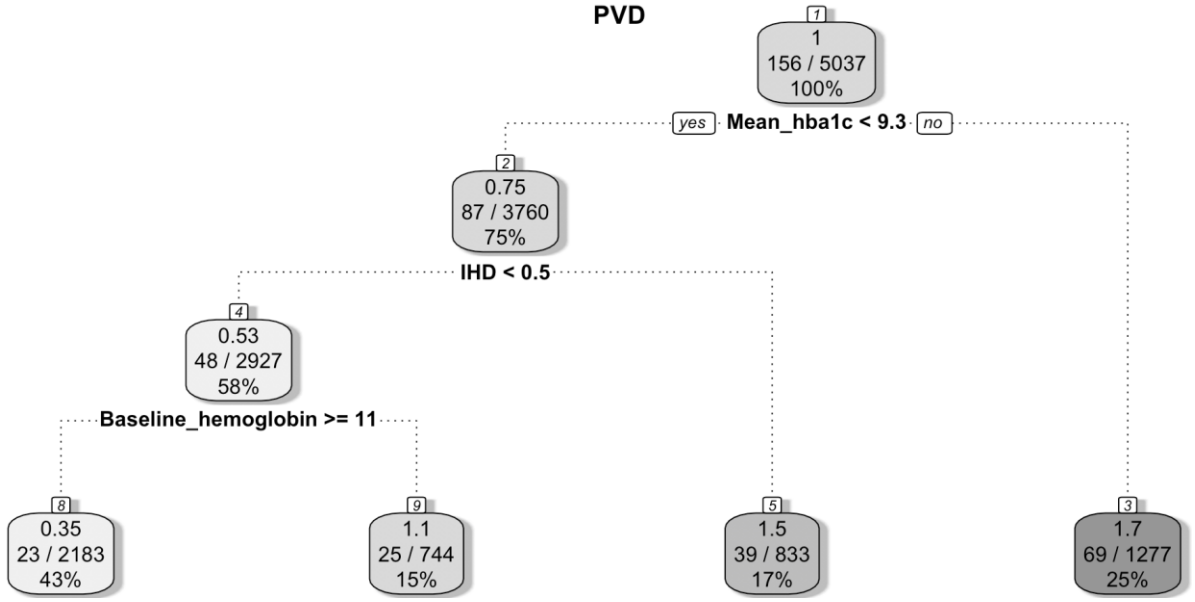
Figure 5. Main tree-based decision rules to predict mortality and complications using regularized and weighted random survival forest model



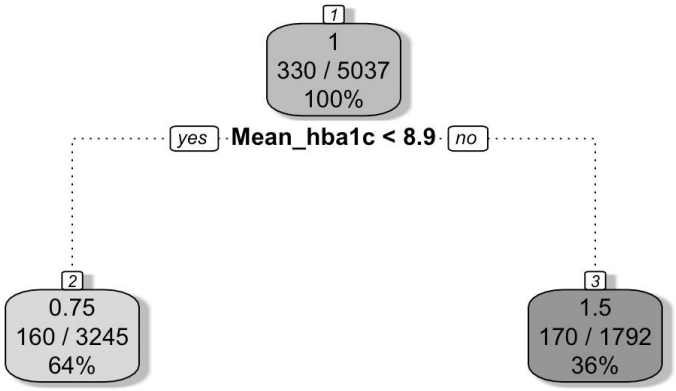
Renal



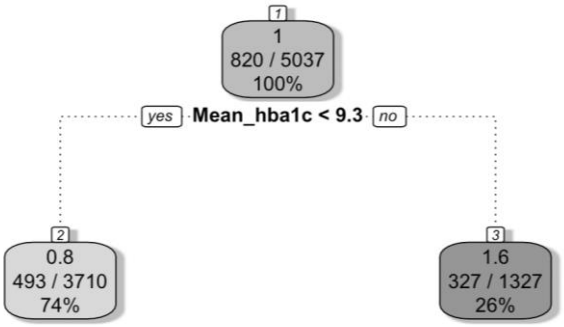
PVD



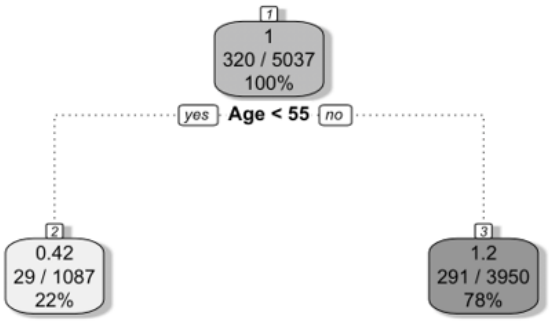
Neurological



Ophthalmological

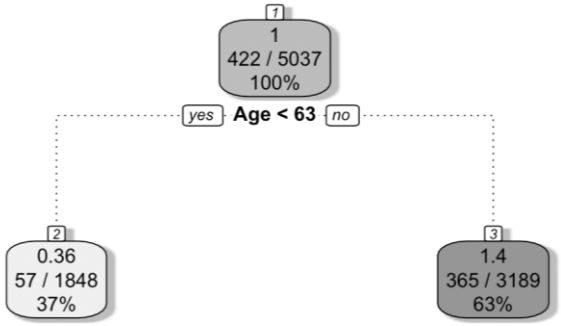


Ischemic Stroke

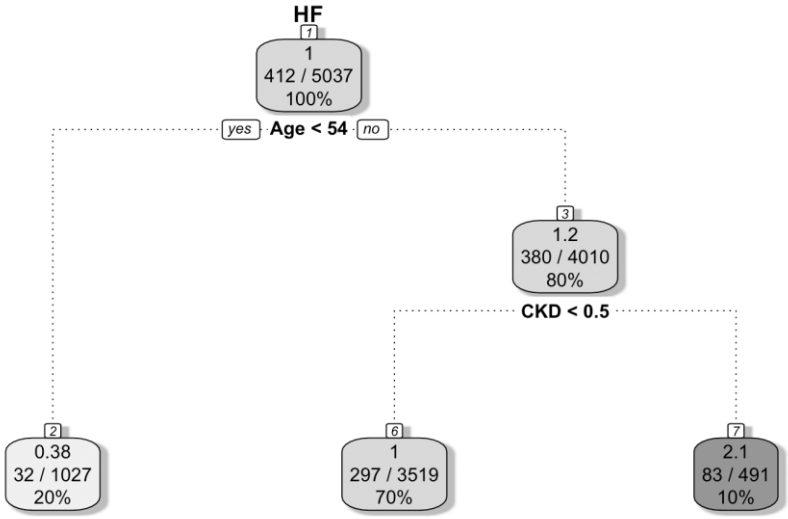


\

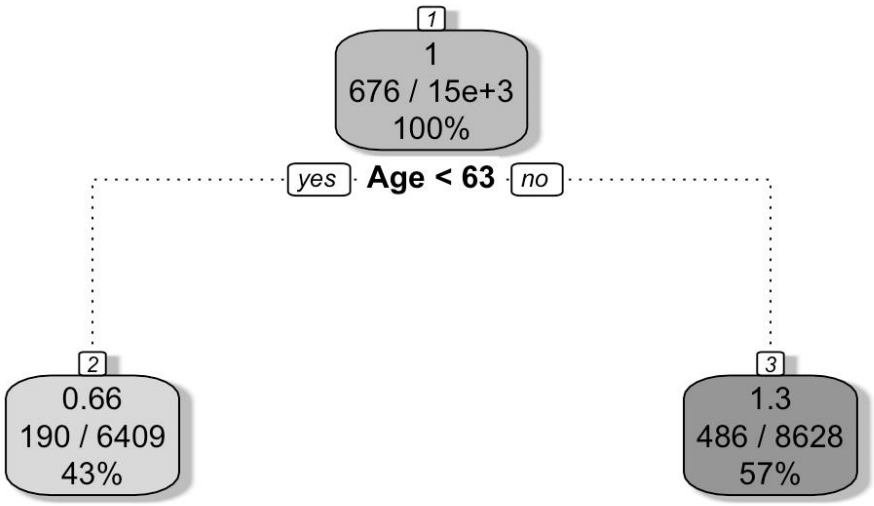
AF

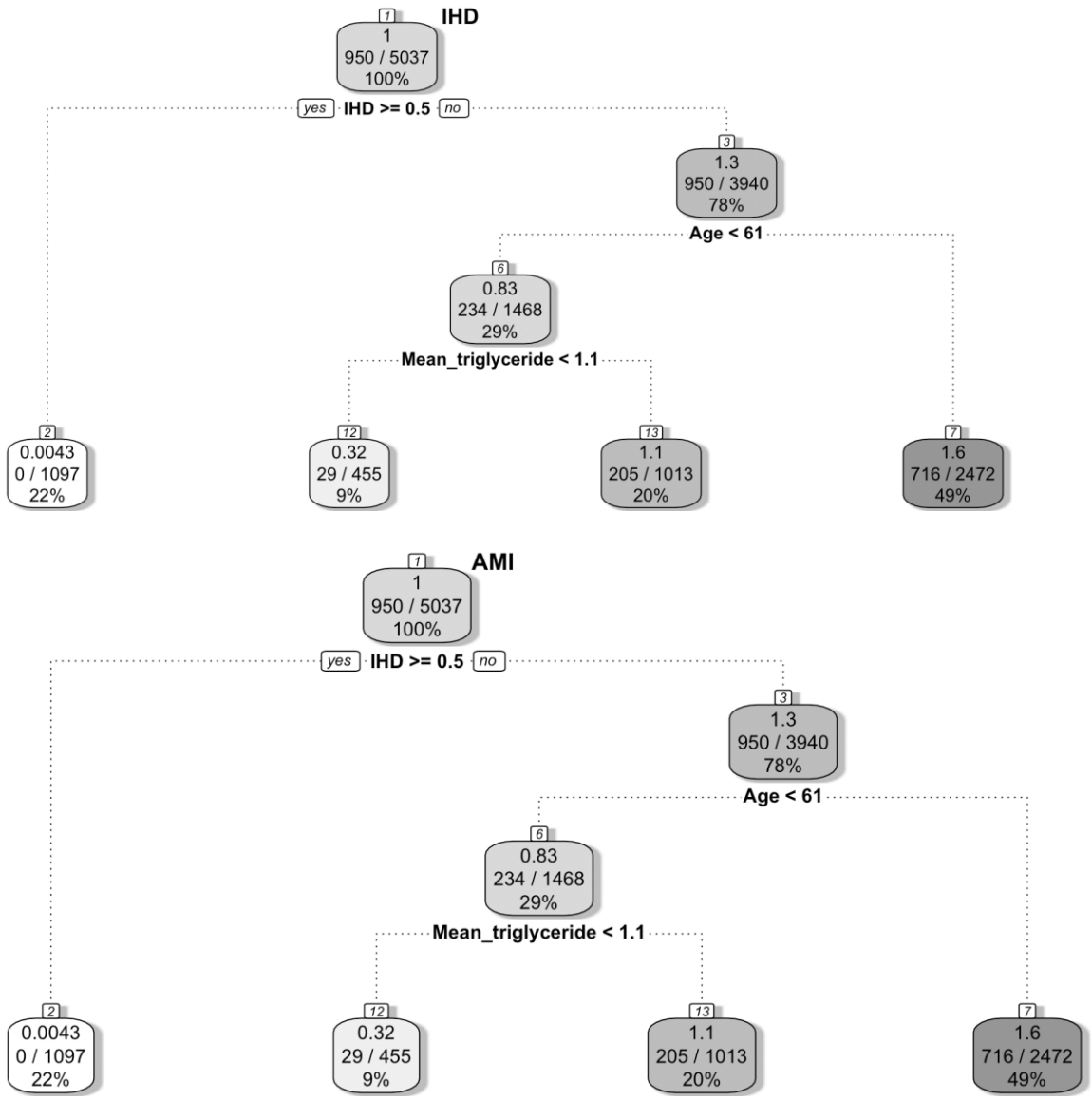

1
1
422 / 5037
100%
yes
Age < 63
no
2
0.36
57 / 1848
37%
3
1.4
365 / 3189
63%

HF


1
1
412 / 5037
100%
yes
Age < 54
no
3
1.2
380 / 4010
80%
CKD < 0.5
2
0.38
32 / 1027
20%
6
1
297 / 3519
70%
7
2.1
83 / 491
10%

ICH


1
1
676 / 15e+3
100%
yes
Age < 63
no
2
0.66
190 / 6409
43%
3
1.3
486 / 8628
57%



CKD: chronic kidney disease; HF: heart failure; AMI: acute myocardial infarction; IHD: ischaemic heart disease; AF: atrial fibrillation; ICH: intracranial haemorrhage

Figure 6 shows the decision tree incorporating variables most important for the prediction of respective outcomes.

The performance of the model for survival analysis of each complication outcome is compared with baselines including RSF and Cox models, based on a five-fold cross-validation approach (**Table 11**). According to the evaluation metric of Harrell's C-index, our model outperforms both RSF and Cox for survival analysis of all-cause mortality, renal complications, PVD, ischemic stroke, AF, HF, ICH, IHD, AMI, and osteoporosis complications, and almost

the same for dementia, neurological, ophthalmological, and complications. The model also shows higher prediction accuracy according to evaluation metrics of precision, recall, and AUC.

Table 11. Model performance comparison analyses with five-fold cross validation

	Our model				RSF				Cox			
	Precision	**Recall**	**AUC**	**C-index**	**Precision**	**Recall**	**AUC**	**C-index**	**Precision**	**Recall**	**AUC**	**C-index**
Mortality	0.9212	0.8663	0.8986	**0.8804**	0.8468	0.8962	0.8377	0.8178	0.7576	0.8025	0.7221	0.7676
Renal	0.9237	0.9180	0.8763	**0.8269**	0.8855	0.8563	0.8577	0.8194	0.7625	0.7803	0.8008	0.7470
PVD	0.8913	0.8565	0.8880	**0.8701**	0.8922	0.8617	0.8517	0.7701	0.7779	0.7289	0.7517	0.7848
Neurological	0.9104	0.9252	0.9111	0.8318	0.8842	0.8223	0.8480	**0.8511**	0.7874	0.7434	0.7969	0.7706
Ophthalmological	0.8902	0.8766	0.9065	0.8208	0.8557	0.8234	0.8643	**0.8237**	0.7671	0.7522	0.7814	0.7517
Ischemic stroke	0.8998	0.9010	0.8885	**0.8527**	0.8367	0.8983	0.8634	0.8484	0.7925	0.7752	0.7884	0.7538
AF	0.9352	0.8641	0.8998	**0.8740**	0.8194	0.8733	0.8523	0.8125	0.7747	0.7571	0.7742	0.7647
HF	0.8963	0.9175	0.8947	**0.8943**	0.8700	0.8533	0.8330	0.7767	0.8047	0.7708	0.7585	0.7749
ICH	0.7893	0.7992	0.7156	**0.7154**	0.7918	0.7796	0.7034	0.7070	0.6405	0.6590	0.6857	0.6414
IHD	0.8829	0.8948	0.9108	**0.8528**	0.8775	0.8375	0.8328	0.87964	0.7720	0.7579	0.7985	0.7738
AMI	0.9073	0.9077	0.8861	**0.8386**	0.8689	0.8455	0.8246	0.7782	0.7815	0.7640	0.7499	0.7845
Osteoporosis	0.7341	0.7142	0.7014	**0.6372**	0.6565	0.7487	0.7372	0.6244	0.6760	0.6717	0.6890	0.5857
Dementia	0.8837	0.8651	0.8660	0.8784	0.8354	0.8594	0.8549	**0.8790**	0.7566	0.7575	0.7345	0.7772

3.4. Discussion and limitations

3.4.1 Discussion

To the best of our knowledge, present study is the first to use RSF to predict for complications and mortality amongst patients with T2DM. There are several major findings of the present study: 1) HbA1c and lipid variability can be used to evaluate the risk for a diverse range of adverse outcomes in diabetes; 2) HbA1c variability is positively associated with increased NLR and frequency of hypoglycaemia episodes; 3) there are interactions present between the value and variability of different lipid parameters.

Although HbA1c and lipid indices were assumed to show a positive linear correlation with mortality risk, there is emerging evidence suggesting that the mortality risk increases at the extreme ends of the parameters. Currie *et al*. first demonstrated the increase in cardiovascular event incidence and all-cause mortality under both low and high mean HbA1c

in 2010, which explained the increased mortality under aggressive glycaemic control in clinical trials. (4, 71) Subsequent cohort studies provided further evidence for the J-shaped association between mean HbA1c and all-cause mortality. (6, 7, 112) Furthermore, recent studies have found that similar to HbA1c, a U-shaped relationship is demonstrated between the lipid indices and adverse outcomes. (113, 114, 115) These findings explain the "reverse epidemiology" observed in both the present study and existing studies, where risk factors for the outcome lower the event risk instead, such as the lowering of intracranial haemorrhage and AF risk under raised mean LDL-C in this cohort. (116) Overall, the J-shaped associations justify the heterogeneous predictions by mean HbA1c and lipid indices.

Interestingly, both the mean and variability of LDL-C were not significant predictor for all-cause mortality after adjusting to the use of lipid-lowering agents in the present study. This may be explained by the introduction of lipid-lowering therapy to patients with high LDL-C levels, which has been shown to lower the risk of MACE (117). Also, it has been reported that the use of statin reduces all-cause mortality over three years amongst patients with T2DM above the age of 65 in the community, independent of multidimensional impairments and age (118). The addition of other lipid-lowering agents, such as ezetimibe, lowers the risk of AMI, ischemic stroke and cardiovascular mortality (119). In addition, patients with higher LDL-C levels may have poorer glycaemic control and a higher cardiovascular risk profile overall, resulting in the use of more antidiabetic agents (120). The protective effects of the antidiabetic agents may have masked the effects of high LDL-C on MACE. The resultant increase in LDL-C variability amongst patients with lowered LDL-C level after medication use may have contributed to the statistical insignificance of LDL-C variability in the prediction for mortality.

Heterogeneity is also demonstrated in the prediction findings of HDL-C variability. Currently, research on the predictive value of HDL-C variability is limited and yields conflicting findings. Whilst some studies report a greater risk for adverse events under

increased HDL-C variability, others reported insignificant findings. (121, 122, 123, 124, 125) Furthermore, as suggested by prior studies, the reflection of lifestyle changes by HDL-C variability may be a contributing factor, where the difference in the effect of interaction between lifestyle factors such as smoking, alcoholism, and physical activity leads to the varied predictive value of HDL-C variability across different outcomes. (125, 126) Since SD is positively correlated to the mean, given the value and variability of HDL-C yields opposite effects, the effects of variability may be reduced when SD is used as a measure of variability. (121) The standardization of variability measures can encourage the application of parameters of variability into clinical practice.

Although the mechanism behind HbA1c and lipid variability is unclear, several hypotheses were raised and explored. Large-scale cohort studies have demonstrated the association between HbA1c variability with all-cause mortality and other adverse outcomes. (60, 83, 127) In terms of HbA1c variability, it is proposed that its relationship to intermittent hypoglycaemia underlies the increased mortality risk. Indeed, our team recently reported a significant relationship between the frequency of hypoglycaemia episodes and HbA1c variability, with the latter predicting all-cause mortality, cardiovascular-specific mortality and various diabetic-related complications. (128) Besides mortality due to hypoglycaemia, a common and lethal complication in diabetes, intermittent hypoglycaemia induces increased oxidative stress (81, 82), causing endothelial dysfunction and chronic inflammation, ultimately leading to increased mortality risk. (79, 80, 129) It has been reported that both acute and chronic glycaemic variability can induce oxidative stress and lead to chronic inflammation. (130) Indeed, increased metabolic variability can induce damage to different organs, leading to complications such as HF. (131) The present study provides supporting evidence for the hypothesis by demonstrating a significant association between HbA1c variability, hypoglycaemic frequency, and baseline NLR. Other than NLR, further inflammatory markers

such as C-reactive protein were found to be associated with HbA1c variability. (132) Similar to HbA1c, the mechanism for lipid variability to increase mortality risk is speculated to be associated with induced oxidative stress. It is speculated that large fluctuations in both LDL-C and HDL-C can lead to plaque instability, therefore releasing atherogenic substances and hence increasing mortality risk. (99, 133) The significant association between baseline NLR and variability across different lipid indices provide insights towards the proposed underlying mechanisms between lipid variability and chronic inflammation. Additionally, the increased variability across biomarkers may reflect generalized frailty. (99)

The effects of anti-diabetic agents on the risk of adverse events in diabetic patients are well studied. (134) In agreement with the present study, sulphonylurea use has been reported to raise the risk of mortality, cardiovascular events, and renal impairment significantly. (135, 136, 137) It should be noted that the use of add-on therapy to insulin may indicate more severe diabetes or used to slow the progression of complications. Hence drug use is the effect, rather than the cause of the adverse outcome. This may explain the increased ophthalmological complication and cardiovascular event risk in biguanide and alpha-glucosidase inhibitors in the present study, contrary to the cardiovascular protective effects reported by existing studies. (138, 139, 140) Additionally, the insignificant effect of DPP4I and thiazolidinedione may be attributed to the fewer number of patients prescribed these drugs in the present cohort. Previously, thiazolidinediones have been associated with a greater risk of HF. In our study, this was associated with a lower risk of HF on univariate Cox regression, but not after propensity score matching for other antidiabetic drugs as seen in other studies from our team on the Hong Kong population(22). Nevertheless, thiazolidinedione has been associated with beneficial effects such as reducing the incidence of AF (141), which are explicable by reverse remodelling. (142, 143, 144, 145) Finally, the annualized mortality rate in our study was 5.87% in our cohort, compared to 1.92% in another local study (57). The reason is that our study cohort included

only diabetic patients who received insulin therapy, which would invariably include those at the highest risk. Moreover, the inclusion of patients who were already on insulin therapy in 2009 meant that few patients benefited from newer anti-diabetic drug classes such as SGLT2I, which can reduce mortality, (146).

Statistical methods such as classification and regression trees are commonly used and are familiar to clinicians but are limited by high variance and poor performance (103, 147). These can be overcome by RSF, which builds hundreds of tree branches and outputs the results by voting (109). RSF reduces variance and bias by using all the collected variables, and then automatically assesses the nonlinear effects and complex interactions amongst them (108). RSF is fully non-parametric, including the effects of the treatments and predictor variables, whereas traditional methods such as the Cox model utilize a linear combination of attributes (148). RSF has been applied in serval risk stratification models for different diseases (149, 150, 151, 152, 153, 154, 155), and has been shown to outperform classical statistical methods, such as the Cox proportional hazards models (104, 149). It should be noted that in the present study, there are HRs close to one, which may indicate a statistically significant, but marginally clinically relevant factor. The large sample size with small portion of patients with the clinical outcome evaluated in the present study may be the underlying reason. Despite the HR being close to one, the marginally significant risk factors in accumulation can still result in clinical significance difference, such as in the cases of polygenic mutations resulting in diseases.

Our study demonstrates the principle that machine learning algorithms can further improve risk prediction of time-to-event (mortality and complications) in diabetic patients receiving insulin therapy. The generated importance rankings and minimal depths of prognostic risk variables can be applied in clinical practice as an easy-for-use complication score for early survival risk identification. Through complication-specific risk stratification amongst diabetic

patients, a personalised management approach with close monitoring for specific complications that individual patients are at high risk of can be adopted.

3.4.2. Strengths and limitations

The major strengths of the present study include: 1) the effects of clinical and biochemical parameters on adverse effects were assessed both independently, and under multivariate analysis each other; 2) the risk for a diverse range of adverse events in diabetes is evaluated; 3) interrelations between chronic inflammation and both HbA1c and lipid variability is explored to give insights on the underlying mechanisms in the pathogenesis; 4) variability is examined by more than one measure to limit the effects of inherent bias; 5) long follow-up period allows for the capture of serial variability and long term adverse outcome.

Several limitations should be noted for the present study. Firstly, similar to other observational studies, there is potential under-coding, missing data, and coding errors. As a result of the missing data across different variables, there is only 7 913 included in the multivariate Cox regression model, which significantly reduces the statistical power of the analysis. Moreover, observational studies can only establish correlation, not causation. Furthermore, the duration of diabetes was not accounted for. However, given that all patients in the study cohort were prescribed insulin for glycaemic control, an advanced stage of diabetes can be inferred. Moreover, there is a large change in the management guidelines, therapeutic options, and treatment targets throughout follow-up. Additionally, there is a lack of data on the patient's BMI and lifestyle factors, such as smoking, alcoholism, and diet, from the database. These variables may affect the lipid levels, in particular HDL-C. The analysis of all-cause mortality is especially affected, given the wide range of contributing factors and influential effect of lifestyle choices. Finally, as the main aim of this study was to examine the predictive values of HbA1c or lipid variability for adverse outcomes, the initial analyses on the

relationships between these variability indices, NLR and hypoglycaemia were exploratory. The inter-relationships between these variables, including the use of mediation analysis, will be explored in future studies. There is also an extensive analysis of adverse outcomes in **Table 9**, which may be difficult to read.

3.5. Conclusion

In conclusion, the present study demonstrates that high HbA1c and lipid variability are associated with an increased risk for adverse outcomes in diabetes across different organ systems. The association between hypoglycaemic frequency and baseline NLR with HbA1c and lipid variability suggests that intermittent hypoglycaemia and chronic inflammation contribute to the mechanism underlying the pathogenic effect of fluctuating glycated haemoglobin and lipid levels. Machine learning techniques, such as RSF, have been incorporated and was able to improve the accuracy of the predictive models. Future studies on the interactions between lipid variability can help to facilitate the application of variability measures in clinical risk stratification. The effects of the sequence of diabetic adverse outcomes on the ultimate patient survival can be explored to gain insights into the systemic pathogenesis of diabetes.

Chapter 4. Development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong

4.1. Introduction

T2DM is one of the most common metabolic conditions, with an increasing prevalence attributable to ageing, sedentary lifestyles, environmental changes, and better disease management. Patients with this condition are at an increased risk of premature death. Existing risk models have been developed, such as QDiabetes for predicting new-onset diabetes (156), in addition to CORE (157), BRAVO (158) and Michigan (159) models for predicting disease progression, complications and mortality. These have generated good predictive results in Western cohorts but are limited by their direct applicability to Asian populations. For example, Chinese patients have a lower body mass index threshold for diabetes development and have a higher propensity to suffer from CKD as a result (160, 161). Whilst Asian population-specific models are available (162, 163, 164, 165), these have generally not incorporated temporal measures of variability for longitudinal data or machine learning approaches, both of which can enhance risk prediction (166, 167). Indeed, with the rapid development of big data analytics, it has become easier to improve discrimination by analysing complex interactions among variables. Previously, a machine learning-driven approach has demonstrated superior performance in predicting diabetes onset in a Chinese cohort (168).

In this territory-wide study, with the aid of machine/deep learning approaches, we developed a risk model for mortality prediction using multi-parametric data from different domains. These include baseline comorbidities, measures of variability of fasting glucose and HbA1c, inflammatory and nutritional indices and drug prescription details. We tested the hypothesis that machine learning methods (RSF (108)) and deep neural survival learning models (DeepSurv (169)) can significantly improve predictive performance when compared to Cox regression-based models.

4.2. Methods

4.2.1. Study design and data source

The study was approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee. The inclusion criteria were patients who received anti-diabetic medications or had ICD-9 codes for T2DM, and attended any of the 43 public hospitals or their associated ambulatory or outpatient facilities managed by the Hong Kong Hospital Authority between January 1st to December 31st, 2009. CDARS was used in this study. This system has been used for epidemiological research by multiple research teams, including our team, in the past (67, 68, 170).

4.2.2. Data extraction

Baseline patient characteristics, including demographic details such as age and sex, prior comorbidities (HF, IHD, ischemic stroke, aborted SCD of all-cause, AMI, AF, PVD, ICH, osteoporosis, dementia, hypertension, COPD, cancer, renal and ophthalmological diabetic complications), anti-diabetic and cardiovascular medications. The duration of living with T2DM from the point of diagnosis till December 31st, 2009 was also extracted, and determined by the earliest fulfilment of any of the following criteria in this order: 1) initial documentation of T2DM-related ICD-9 codes; 2) earliest HbA1c >6.5%; 3) earliest FBG > 7 mmol/L. Time-till all-cause mortality was determined as the number of days from the starting date of patient inclusion, January 1st, 2009, till the day of death or the end of the follow up period, December 31st, 2019.

The following laboratory data were collected at baseline: NLR was derived by dividing the absolute neutrophil by the lymphocyte count, anaemia defined as <13g/dL for males and <12g/dL for females, biochemical test results including 1) creatinine, sodium, potassium, 2)

urea, 3) albumin and total protein, 4) ALT and ALP, 5) FBG and HbA1c; 6) HDL-C, LDL-C, total cholesterol, and triglyceride.

The number of anti-diabetic drugs by class was extracted: 1) insulin, 2) biguanide, 3) sulphonylurea, 4) alpha-glucosidase inhibitor, 5) thiazolidinedione, 6) DPP4I, 7) GLP-1A, 8) meglitinide. Similarly, the number of anti-hypertensive medications of the following classes were also extracted: 1) ACEI/ ARB, 2) beta-adrenergic receptor blocker, 3) CCB, 4) diuretics. Lipid-lowering agents were also extracted.

4.2.3. Variability calculations

To calculate FBG and HbA1c variability, data points were obtained for the period between January 1st, 2004 and December 31st, 2008. Only patients with three or more measurements for the specific parameter were included in the variability analysis of the respective parameter. The different measures are detailed below.

1) SD, 2) absolute variability score defined as 100 x no. of measurements > 0.5 / no. of measurements, 3) percentage variability score defined as 100 x no. of measurements > 10% of previous measurement/ no. of measurements, 4) normalized absolute variability score given by (2) / individual mean, 5) normalized percentage variability score given by (3) / individual mean, 6) SD/ individual baseline, 7) coefficient of variation given by SD / individual mean, 8) variability independent of mean given by SD/ individual mean^(ln(population SD)/ln(population mean)).

4.2.3. Outcomes and statistical analysis

The primary outcome for the present study is all-cause mortality. Univariate Cox regression was applied to identify significant predictors for all-cause mortality and HR with 95% CI were reported. Variables achieving p <0.10 were included in a diabetes duration-adjusted multivariate model to improve the chances of including variables that are significant only in the presence of others (171). Statistical significance is defined as P-value <0.05. FBG and HbA1c variability of the same formula were paired and added to the multivariate model to assess their predictiveness through comparison of HR.

To generate a predictive score, Cox regression was repeated for the final multivariate model with measures of variability included. HR between 1-1.50 were awarded 1 mark in the score. To adjust for the U-shaped relationship against mortality reported for HDL-C, LDL-C, total cholesterol and HbA1c, these parameters were first divided by deciles to serve as cut-offs and undergo univariate Cox regression. Thereafter, the decile with the smallest HR was selected as a reference and compared against the remaining deciles through univariate Cox regression again. The minimum and maximum cut-offs for the deciles that had insignificant differences with the reference decile were selected as the cut-offs to be used in the score. To demonstrate the U-shaped relationship, the HR of deciles were plotted graphically. Cut-off values for continuous variables in the score were found through maximizing sensitivity and specificity. Age and diabetes duration were rounded to the nearest whole number, whilst other parameters were rounded to two decimal points. The predictive value of the score was evaluated through the generation of a ROC curve and AUC calculated.

To further evaluate the predictive value of the measures of variability, the measures were also divided into quartiles, with the first quartile as reference, to perform univariate Cox regression and assess the AUC of the quartile cut-offs. The quartile HR of the FBG and HbA1c

measures of variability were illustrated graphically. Statistical analyses were performed using RStudio software (Version: 1.1.456) and Python (Version: 3.6).

4.2.4. Development of machine/ deep models for survival learning

Machine/deep learning survival analysis models can directly capture the relationships between risk predictors and mortality outcomes without prior functional assumptions typically made in Cox analysis models. Here we used an RSF model, a type of machine learning method for survival analysis, relying on the intuition that the best survival learning model, when combined with weak decision tree learning models, can minimize the overall survival prediction errors. The prediction errors are measured by performance evaluators, e.g., precision, recall, AUC, and C-index. The OOB method was adopted whenever a bootstrap sample (bag ones) was down with a replacement from the training dataset. The bootstrapping technique is used to grow the tree and results in well-defined subsets. Some of the bootstraps are duplicates and are members of the in-bag subset, and the remaining individuals define the OOB subset for the final tree. Each individual in the OOB subset for a tree is passive. A unique terminal node membership and terminal node statistic were assigned. An OOB ensemble statistic for each individual is formed by combining the terminal node statistics from all trees where an individual is an OOB member. Finally, the class with the maximum frequency in the OOB ensemble statistic serves as the predicted class label for the member.

The variable's importance of interest is calculated as the prediction error (squared loss) of the original ensemble event-specific cumulative probability function subtracted from the prediction error of the original ensemble event-specific cumulative probability function (obtained when each OOB instance is just dropped down its in-bag competing risks tree) (110) In this study, RSF was used for mortality prediction and the most important predictors were

ranked according to variable importance measure in RSF. Variables that were important predictors of risk outcome have a larger importance value, indicating higher predictive strength, whereas non-predictive variables have zero or negative values.

We further employed a nonlinear deep learning survival method termed the Cox proportional hazards DeepSurv approach. This can inherently and adaptively model the high-level interaction patterns among risk predictors and thus can better capture the complex nonlinear relationship between patients' covariates (e.g. clinical features) and mortality outcome directly. In contrast, standard survival models like the linear Cox proportional hazards model require extensive feature engineering and necessary prior medical knowledge to model mortality risk at an individual level. Specifically, DeepSurv is a deep feed-forward neural network that can predict the effects of a patient's baseline covariates on their hazard rate parameterized by the weights of the neural network. The input of DeepSurv is the baseline variable of the diabetic patient. The hidden layers of DeepSurv consist of a fully connected layer of nodes, followed by a dropout layer(172). The output of the DeepSurv is a single node with a linear activation which estimates the log-risk function in the Cox model. In this study, we train DeepSurv by presetting the objective function to be the average negative log form of Cox partial likelihood with L2-regularization (173), to model for mortality risk prediction of diabetic patients. Gradient descent optimization was used to find the weights of DeepSurv. The hyper-parameters of DeepSurv including the number of hidden layers, number of nodes in each layer, and dropout probability were determined from a random hyper-parameter search approach (174).

A five-fold cross validation approach was performed to compare the survival prediction performance of RSF and DeepSurv in terms of precision, recall, AUC, and C-index over the standard Cox model. The R packages, *randomForestSRC* (Version 2.9.3), *ggplot2* (Version

3.3.2), and python package *DeepSurv* (Version 0.1.0) were used to generate the mortality prediction results.

4.3. Results

4.3.1. Baseline characteristics

The study cohort included 273876 patients (mean age: 65.4 ± 12.7 years, male: 48.2%, diabetes duration= 6.18 ± 4.56 years) with a median follow-up of 142 (interquartile range (IQR)= 106-142) months, which corresponded to a total of 2660465 patient-years. The baseline demographics, clinical, laboratory and drug details are shown in **Tables 12 and 13** for continuous and discrete variables, respectively. The most prevalent comorbidities were hypertension, IHD and HF. The percentage of patients on n=0, 1, 2, 3, and 4 anti-diabetic medications were 13.3%, 34.8%, 46.1%, 5.4%, and 0.4% respectively. At baseline, the fasting glucose and HbA1c were 8.02 ± 1.95 mmol/L and 7.75 ± 2.59% respectively. The median number for fasting glucose and HbA1c measurements were 7 (IQR= 4-11) and 7 (IQR=4-10) respectively. The different measures of variability for fasting glucose or HbA1c are quantified for subsequent use to predict mortality.

Table 12. Baseline characteristics for continuous variables for the development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong

Characteristics	Mean	Standard Deviation
Age	65.4	12.7
Follow-up Duration (days)	3546	1208
Diabetes Duration (years)	6.18	4.56
Liver Function Test		

Alkaline Phosphatase (U/L)	81.1	37.6
Alanine Aminotransferase (U/L)	28.8	52.9
Total Protein (g/L)	74.5	6.67
Albumin (g/L)	38.9	5.04
Complete Blood Count		
Lymphocyte Count ($x10^9$/L)	1.89	1.04
Neutrophil Count ($x10^9$/L)	5.35	2.69
Neutrophil-Lymphocyte Ratio	3.72	4.37
Haemoglobin Count ($x10^9$/L)	12.8	1.86
Lipid Profile		
High Density Lipoprotein Cholesterol (HDL-C) (mmol/L)	1.23	0.348
Low Density Lipoprotein Cholesterol (LDL-C) (mmol/L)	3.09	0.941
Total Cholesterol (mmol/L)	5.12	1.13
Triglyceride (mmol/L)	1.63	1.51
Renal Function Test		
Creatinine (umol/L)	102	87.2
Potassium (mmol/L)	4.24	0.522
Sodium (mmol/L)	139	3.48
Urea (mmol/L)	6.96	4.11
Glycaemic Control		
Fasting Blood Glucose	8.02	1.95
HbA1c	7.75	2.59

Table 13. Baseline characteristics for discrete variables for the development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong

Characteristics	Number	Percentage
Male	132040	48.2

Baseline Anaemia	39799	14.5
Anti-Diabetic Agent		
Biguanide	185881	67.9
Sulphonylurea	173525	63.4
Insulin	29697	10.8
Meglitinide	27	0.01
Dipeptidyl Peptidase-4 Inhibitor	325	0.12
Thiazolidinedione	3698	1.35
Glucagon-like Peptide-1 Agonist	17	0.006
Acarbose	3292	1.20
Cardiovascular Drugs		
Angiotensinogen converting enzyme inhibitor (ACEI)/ angiotensin receptor blocker (ARB)	121786	44.5
Beta-adrenergic receptor blocker	92309	33.7
Calcium Channel Blocker	109225	39.9
Diuretic	52096	19.0
Lipid-Lowering Agent	61401	22.4
Comorbidities		
Diabetic Renal Complication	3381	1.23
Peripheral Vascular Disease (PVD)	346	0.13
Diabetic Ophthalmological Complication	3543	1.29
Ischemic Stroke	8986	3.28
Sudden Cardiac Death (SCD)	6420	2.34
Atrial Fibrillation (AF)	7772	2.84
Heart Failure (HF)	11189	4.09
Intracranial haemorrhage	3264	1.19
Ischemic Heart Disease (IHD)	26423	9.65
Osteoporosis	137	0.050
Dementia	2842	1.04

Hypertension	64246	23.5
Chronic Obstructive Pulmonary Disease	818	0.299
Cancer	12190	4.45

4.3.2. Predictors of all-cause mortality

Over a median follow-up period of 142 (IQR=106-142) months, 91155 deaths were recorded (33.3%), which corresponded to an annualized mortality rate of 3.43%. The significant univariate predictors for all-cause mortality are presented in **Table 14**. All measures of variability for FBG and HbA1c were significant predictors as well. The graphical comparison of HR from quartile cut-offs of FBG and HbA1c variability predictors is shown in **Figures 6 and 7**.

Table 14. Univariate predictors for all-cause mortality in diabetic patients in Hong Kong

	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.090	[1.089, 1.091]	< 0.0001
Male	1.12	[1.11, 1.14]	< 0.0001
Complete Blood Count			
Neutrophil-Lymphocyte Ratio	1.033	[1.032, 1.034]	< 0.0001
Baseline Anaemia	3.50	[3.45, 3.55]	< 0.0001
Lipid Profile			
High Density Lipoprotein Cholesterol (HDL-C)	0.836	[0.815, 0.857]	< 0.0001
Low Density Lipoprotein Cholesterol (LDL-C)	0.883	[0.874, 0.892]	< 0.0001
Total Cholesterol	0.910	[0.903, 0.916]	< 0.0001
Triglyceride	0.963	[0.957, 0.970]	< 0.0001

Comorbidity			
Renal Diabetic Complication	3.68	[3.54, 3.83]	< 0.0001
Ophthalmological Diabetic Complication	2.73	[2.62, 2.84]	< 0.0001
Peripheral Vascular Disease	4.39	[3.91, 4.93]	< 0.0001
Ischemic Stroke	2.85	[2.78, 2.93]	< 0.0001
Sudden Cardiac Death	2.48	[2.40, 2.56]	< 0.0001
Atrial Fibrillation	3.54	[3.45, 3.64]	< 0.0001
Heart Failure	4.74	[4.64, 4.85]	< 0.0001
Intracranial haemorrhage	2.70	[2.59, 2.82]	< 0.0001
Ischemic Heart Disease	2.24	[2.20, 2.28]	< 0.0001
Osteoporosis	2.87	[2.34, 3.52]	< 0.0001
Dementia	5.92	[5.69, 6.16]	< 0.0001
Hypertension	2.55	[2.52, 2.59]	< 0.0001
Chronic Obstructive Pulmonary Disease	4.55	[4.22, 4.91]	< 0.0001
Cancer	2.48	[2.42, 2.54]	< 0.0001
Fasting Blood Glucose (FBG)			
Mean	1.00	[0.997, 1.01]	0.527
Absolute Successive Variability Score	1.008	[1.007, 1.008]	< 0.0001
Percentage Successive Variability Score	1.01	[1.009, 1.01]	< 0.0001
Standard Deviation	1.15	[1.15, 1.16]	< 0.0001
Normalized Absolute Successive Variability Score	1.065	[1.06, 1.07]	< 0.0001
Normalized Percentage Successive Variability Score	1.07	[1.067, 1.074]	< 0.0001
Standard Deviation/ Initial FBG	1.01	[1.009, 1.01]	< 0.0001
Coefficient of Variation	1.019	[1.018, 1.019]	< 0.0001
Variability Independent of Mean	1.011	[1.01, 1.011]	< 0.0001
HbA1c			

Mean	1.07	[1.06, 1.07]	< 0.0001
Absolute Successive Variability Score	1.01	[1.007, 1.01]	< 0.0001
Percentage Successive Variability Score	1.008	[1.08, 1.009]	< 0.0001
Standard Deviation	1.19	[1.18, 1.20]	< 0.0001
Normalized Absolute Successive Variability Score	1.055	[1.05, 1.06]	< 0.0001
Normalized Percentage Successive Variability Score	1.063	[1.06, 1.07]	< 0.0001
Standard Deviation/ Initial HbA1c	1.014	[1.01, 1.014]	< 0.0001
Coefficient of Variation	1.017	[1.016, 1.018]	< 0.0001
Variability Independent of Mean	1.011	[1.01, 1.012]	< 0.0001

Figure 6. Graphical representation of quartile hazard ratios from fasting blood glucose variability measures

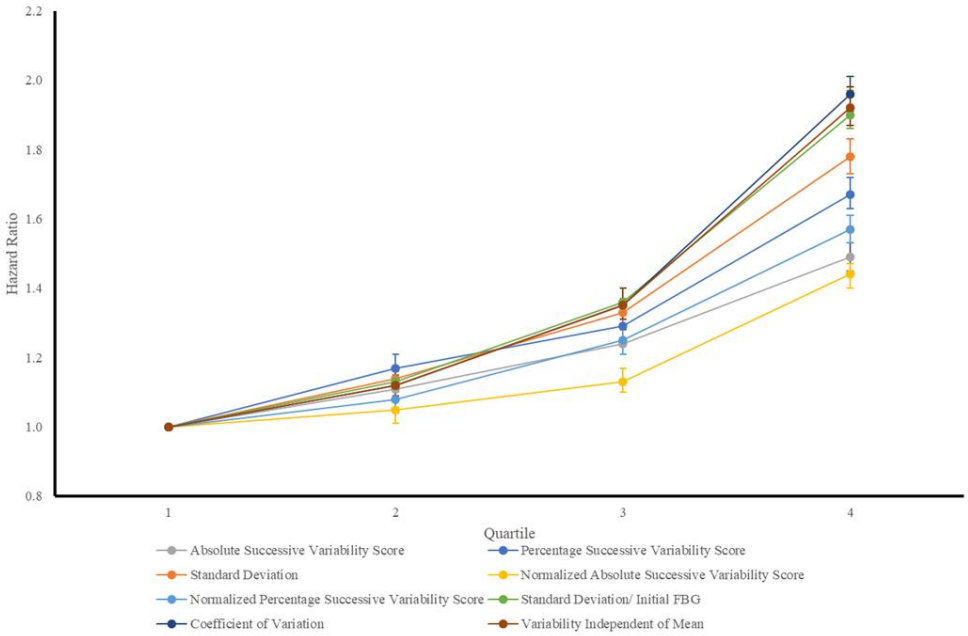
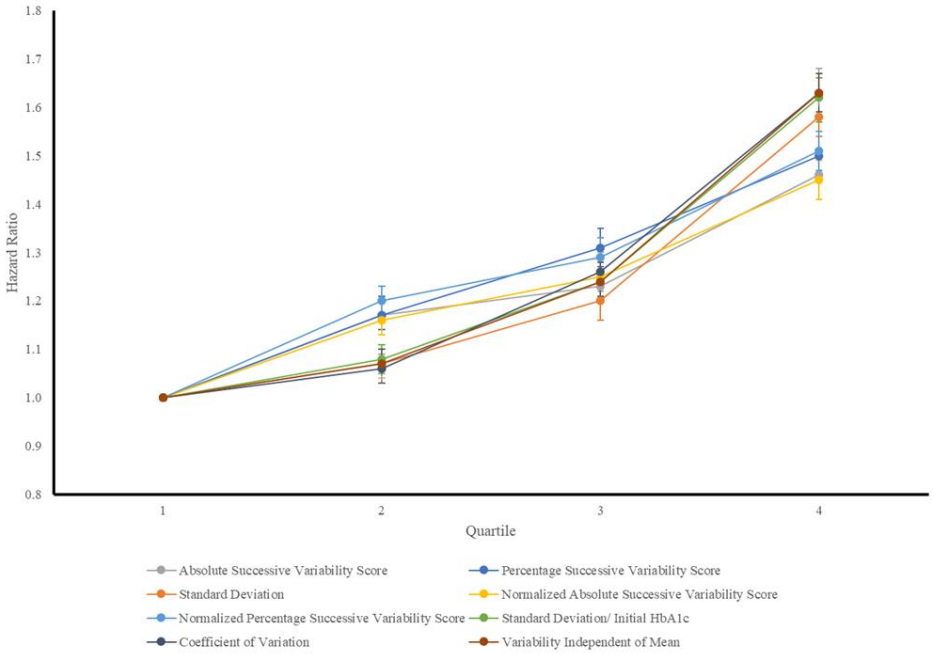


Figure 7. Graphical representation of quartile hazard ratios from HbA1c variability measures



The following parameters remained significant predictors following multivariate adjustment (**Table 15**): 1) age and male gender, baseline comorbidities or complications (hypertension, HF, AF, COPD, cancer, dementia, ischemic stroke, ICH, aborted SCD, diabetic renal and ophthalmological complications), 3) laboratory tests (anaemia, NLR; HDL-C, total cholesterol, triglyceride; mean HbA1c and mean FBG), 4) eight different measures of variability for HbA1c and FBG. A U-shaped relationship between HDL-C, LDL-C, and total cholesterol, but not for triglyceride and all-cause mortality. A U-shaped relationship was also observed for HbA1c but not for FBG (**Figure 8**).

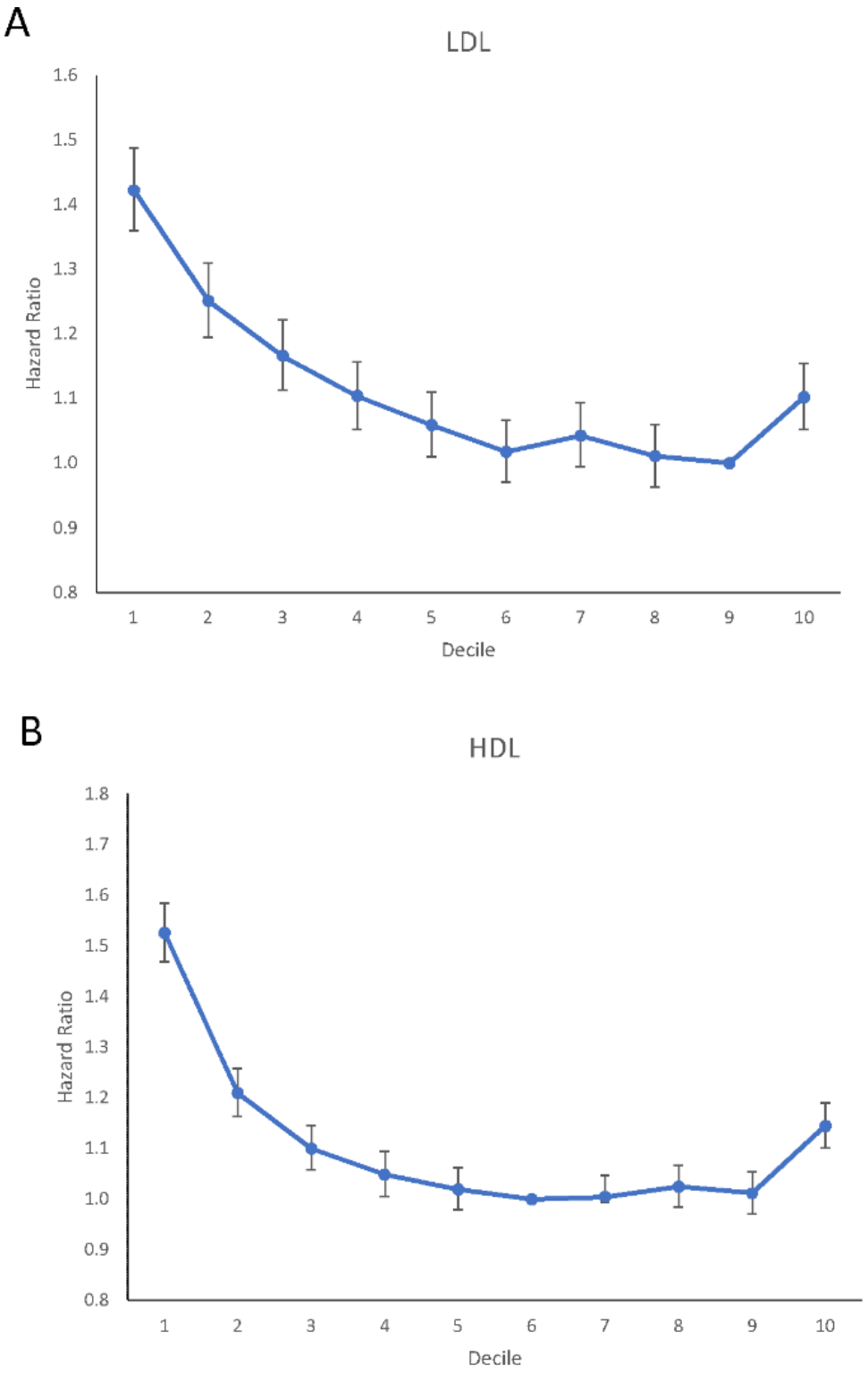
Table 15. Multivariate predictors for all-cause mortality in diabetic patients in Hong Kong

	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.06	[1.06, 1.06]	< 0.0001
Male	1.35	[1.31, 1.40]	< 0.0001

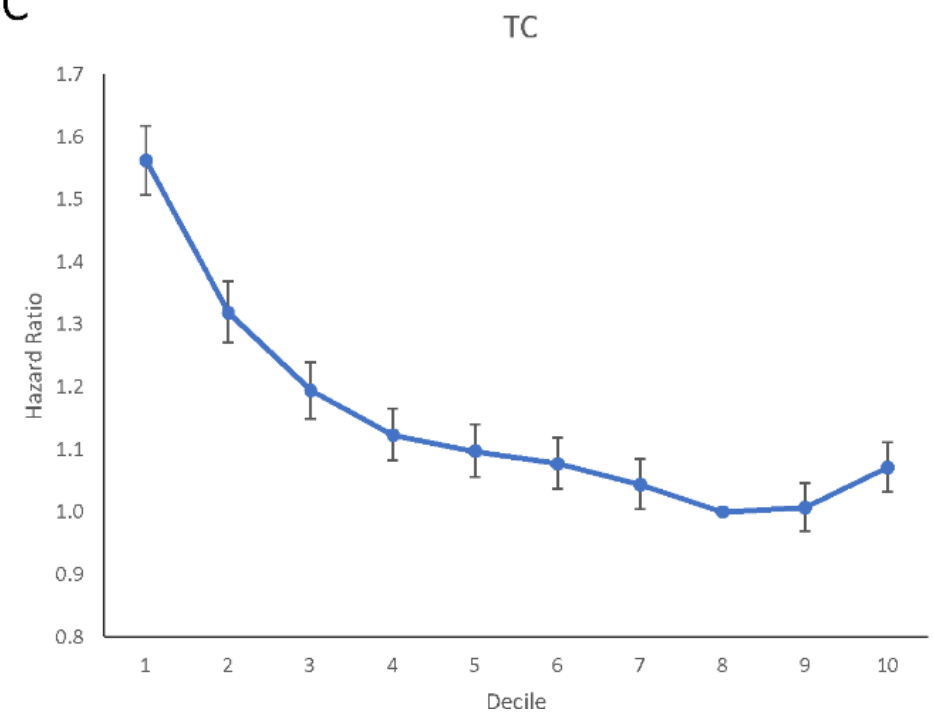
Complete blood count			
Neutrophil-Lymphocyte Ratio	1.02	[1.02, 1.03]	< 0.0001
Baseline Anaemia	1.94	[1.87, 2.01]	< 0.0001
Lipid Profile			
High Density Lipoprotein Cholesterol	0.891	[0.849, 0.935]	< 0.0001
Low Density Lipoprotein Cholesterol	1.01	[0.986, 1.04]	0.348
Total Cholesterol	1.04	[1.01, 1.06]	0.001
Triglyceride	1.02	[1.01, 1.03]	0.001
Comorbidity			
Renal Diabetic Complication	1.28	[1.20, 1.36]	< 0.0001
Ophthalmological Diabetic Complication	1.18	[1.11, 1.26]	< 0.0001
Peripheral Vascular Disease	1.16	[0.984, 1.37]	0.078
Ischemic Stroke	1.25	[1.18, 1.32]	< 0.0001
Sudden Cardiac Death	1.17	[1.09, 1.25]	< 0.0001
Atrial Fibrillation	1.30	[1.23, 1.37]	< 0.0001
Heart Failure	1.62	[1.54, 1.69]	< 0.0001
Intracranial haemorrhage	1.28	[1.16, 1.41]	< 0.0001
Ischemic Heart Disease	1.01	[0.971, 1.05]	0.574
Osteoporosis	1.03	[0.769, 1.38]	0.842
Dementia	1.81	[1.64, 2.00]	< 0.0001
Hypertension	1.30	[1.26, 1.35]	< 0.0001
Chronic Obstructive Pulmonary Disease	1.43	[1.20, 1.70]	< 0.0001
Cancer	1.41	[1.33, 1.49]	< 0.0001
Mean Fasting Blood Glucose (FBG)	1.01	[1.00, 1.02]	0.011
Mean HbA1c	1.06	[1.04, 1.08]	< 0.0001
Fasting Blood Glucose (FBG)			

Absolute Successive Variability Score	1.00	[1.00, 1.00]	0.033
Percentage Successive Variability Score	1.00	[1.00, 1.00]	< 0.0001
Standard Deviation	1.08	[1.07, 1.10]	< 0.0001
Normalized Absolute Successive Variability Score	1.02	[1.01, 1.02]	< 0.0001
Normalized Percentage Successive Variability Score	1.03	[1.02, 1.03]	< 0.0001
Standard Deviation/ Initial	1.00	[1.00, 1.00]	< 0.0001
Coefficient of Variation	1.01	[1.01, 1.01]	< 0.0001
Variability Independent of Mean	1.01	[1.01, 1.01]	< 0.0001
HbA1c			
Absolute Successive Variability Score	1.00	[1.00, 1.00]	< 0.0001
Percentage Successive Variability Score	1.00	[1.00, 1.00]	< 0.0001
Standard Deviation	1.11	[1.07, 1.14]	< 0.0001
Normalized Absolute Successive Variability Score	1.02	[1.01, 1.03]	< 0.0001
Normalized Percentage Successive Variability Score	1.03	[1.02, 1.04]	< 0.0001
Standard Deviation/ Initial	1.01	[1.01, 1.01]	< 0.0001
Coefficient of Variation	1.01	[1.00, 1.01]	< 0.0001
Variability Independent of Mean	1.00	[1.00, 1.01]	< 0.0001

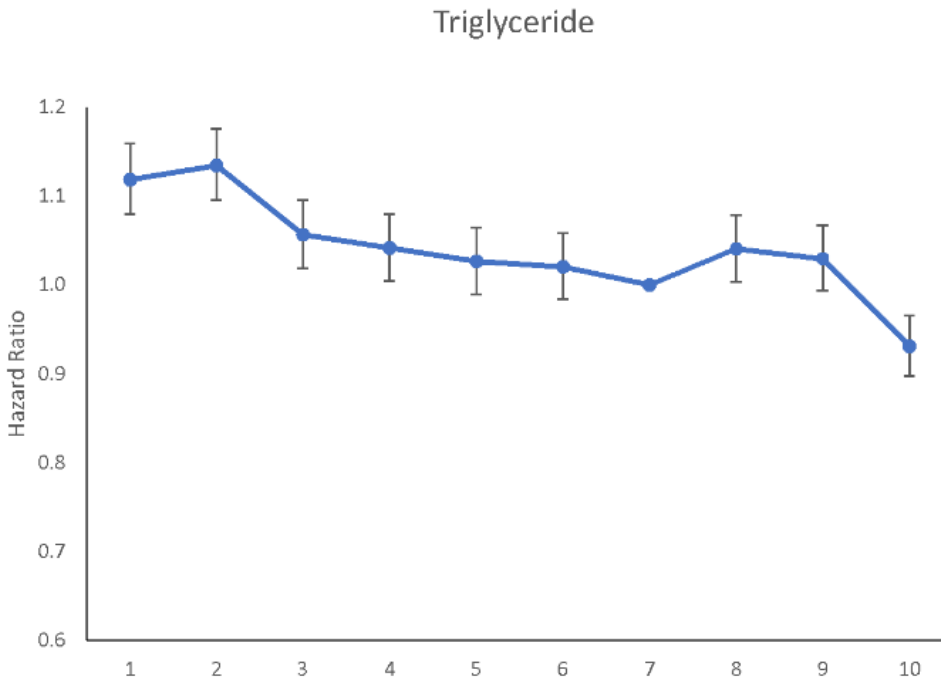
Figure 8. Graphical representation of hazard ratios for all-cause mortality cause mortality



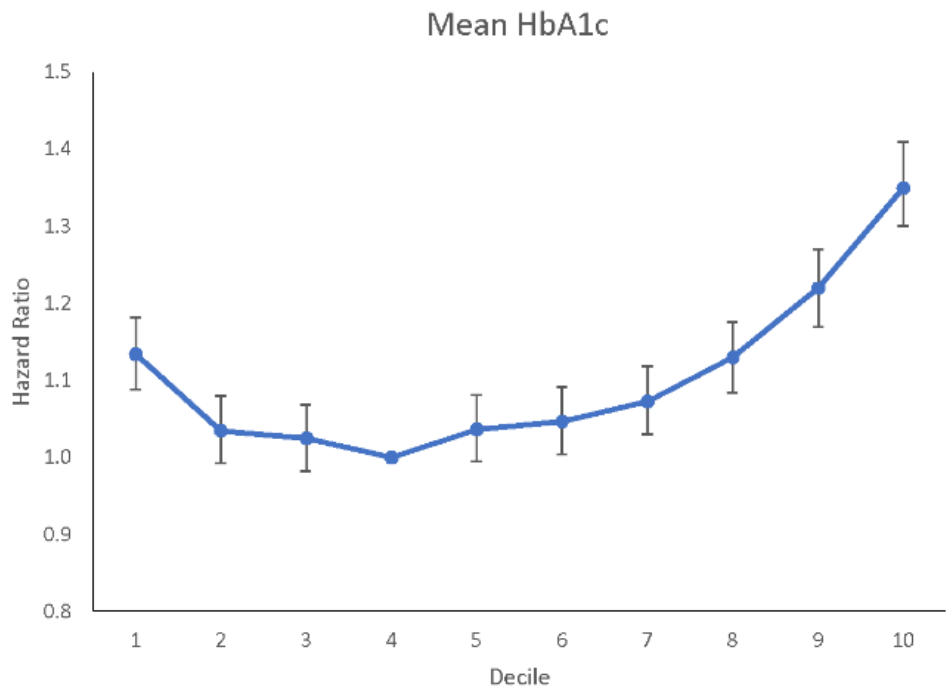
C


TC
Hazard Ratio
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
1
2
3
4
5
6
7
8
9
10
Decile

D


Triglyceride
Hazard Ratio
1.2
1.1
1.0
0.9
0.8
0.7
0.6
1
2
3
4
5
6
7
8
9
10

E



F

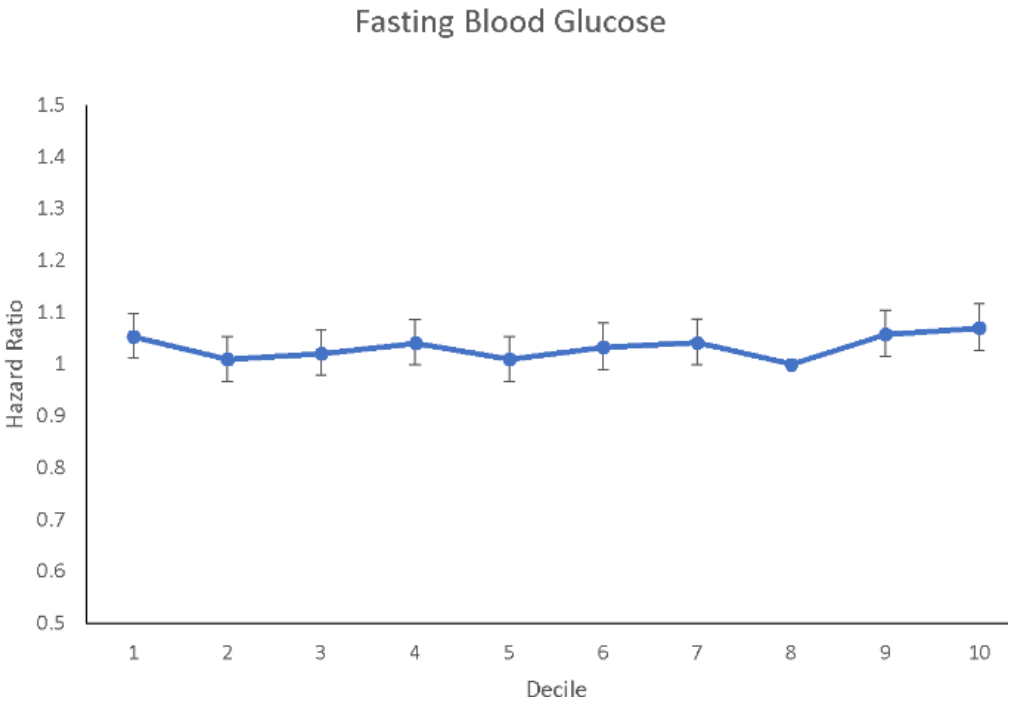


Figure 9: A: low-density lipoprotein-cholesterol (LDL-C); B: high-density lipoprotein-cholesterol (HDL-C), C: total cholesterol; D: Triglyceride. E: mean HbA1c. F: mean fasting blood glucose. The data points were generated by comparing the risks for mortality between patients in decile one and the other respective decile.

4.3.3. Development of a score-based predictive risk model based on Cox regression

A score-based predictive risk model for all-cause mortality was developed by incorporating significant predictors from multivariate analysis. One point was allocated for each significant predictor where the HR was less than 1.5, and 2 points for HRs between 1.5 and 2.5. Out of the eight measures of variability for HbA1c and FBG, SD had the highest HR and greatest statistical significance when adjusted to the multivariate model (FBG: HR= 1.08, 95% CI= [1.07, 1.10], $p < 0.0001$; HbA1c: HR= 1.11, 95% CI= [1.07, 1.14], $p < 0.0001$). It was therefore selected to be included in the mortality score. Altogether, the predictive risk model had a total score out of 25 (**Table 16**). ROC analysis was performed, demonstrating an AUC of 0.729 (**Figure 9**).

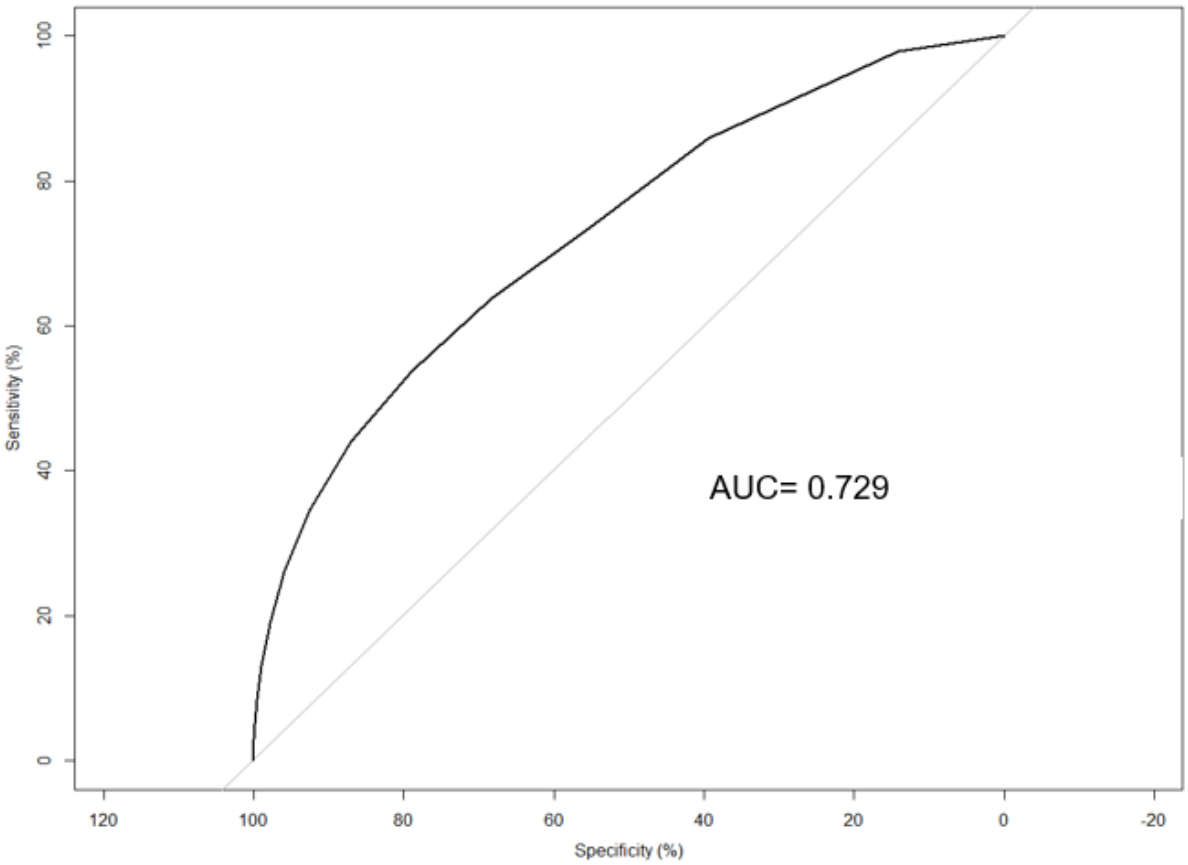
Table 16. A score-based predictive risk model for all-cause mortality in type 2 diabetes mellitus

	Criteria	Score
Age	>70	1
Male	Male	1
Complete Blood Count		
Neutrophil-Lymphocyte Ratio	>2.85	1
Baseline Anaemia	Present	2
Lipid Profile		
High Density Lipoprotein-Cholesterol (mmol/L)	<1.10 or > 1.67	1
Total Cholesterol (mmol/L)	<5.60 or > 6.50	1
Total Triglyceride (mmol/L)	>1.24	1
Comorbidity		
Renal Diabetic Complication	Present	1
Ophthalmological Diabetic Complication	Present	1
Peripheral Vascular Disease	Present	1

Ischemic Stroke	Present	1
Sudden Cardiac Death	Present	1
Atrial Fibrillation	Present	1
Heart Failure	Present	1
Intracranial haemorrhage	Present	1
Dementia	Present	2
Hypertension	Present	1
Chronic Obstructive Pulmonary Disease	Present	1
Cancer	Present	1
Fasting blood glucose and HbA1C: baseline mean and measures of variability		
Mean HbA1c (%)	< 6.34 or > 7.52	1
Mean Fasting Blood Glucose (FBG) (mmol/L)	> 6.12	1
Standard Deviation: FBG	> 1.63	1
Standard Deviation: HbA1c	> 0.79	1

The cut-off for age is rounded to the nearest whole number.

Figure 10. Receiver operating characteristic curve and area under the curve of the score-based predictive risk model for all-cause mortality



4.3.4. Results of machine/ deep learning approaches for risk modelling

We performed the RSF model to predict mortality outcomes based on the variables mentioned in **Table 16**. The optimal tree number of the RSF model was selected as 400 using a five-fold cross validation approach to minimize the overall squared error rate in the testing set as shown in **Figure 10**. RSF model generated the variable importance ranking as shown in **Table 17**, and patient age, prior HF, baseline anaemia, NLR, cancer, hypertension, HDL-C, and renal diabetic complication ranked as the most important predictors to estimate mortality probability, followed by HbA1c measures of variability including HbA1c CV, SD of HbA1c, and FBG measures of variability such as FBG CV, variability independent of mean FBG, absolute successive variability score of FBG. By identifying important variables, patients with

prior risks factors such as pre-existing diabetic nephropathy, can be identified for early intervention. Modifiable risk factors, such as baseline anaemia and hypertension, can be treated to lower the patient's risk.

Figure 11. Selecting optimal tree number for random survival forest model

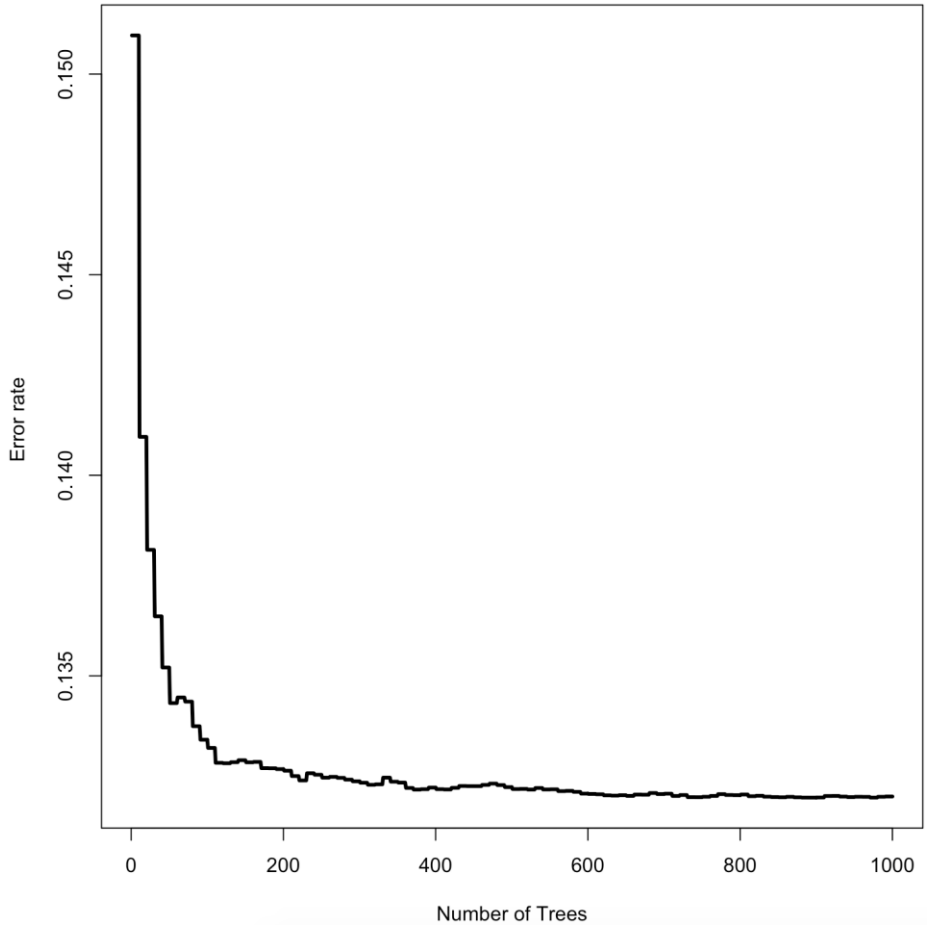


Table 17. Variable importance ranking generated by random survival forest model

Characteristics	**Importance**
Neutrophil-lymphocyte ratio	0.1373
Age	0.0215
Anaemia	0.0167
Standard deviation of fasting blood glucose	0.0112
Heart failure	0.0061
Standard deviation of HbA1c	0.0054
Mean fasting blood glucose	0.0038
Hypertension	0.0030
High density lipoprotein-cholesterol	0.0022

Finally, we compared the survival analysis performance of the RSF model and DeepSurv as typical machine learning and deep learning approaches, respectively, over the multivariate Cox model to predict the mortality outcome of the diabetic patients using five-fold cross validation method. Sobol solver (175) was used to sample each hyper-parameter of DeepSurv from a predefined range and k-means cross validation (k=3) was used to evaluate the performance of the parameter configuration settings. Using the configuration with the largest validation C-index on the testing set to avoid models that overfit, we selected the best hyper-parameters of the DeepSurv network which included: number of dense layers=4, learning rate=0.0003, ℓ2 regularization coefficient=3.25, dropout rate=0.36, exponential learning rate decay constant=0.0005, and momentum=0.86. In all instances, the ReLU activation function was applied. (176).

The comparative performance results of the different models are shown in **Table 18.**It should be noted that although both RSF and DeepSurv are machine learning models, it would be difficult to compare the models side by side since DeepSurv operates as a neural network, where the sequence of variables inputted would have an effect on the predictive performance. However, both models worked with the same set of variables. Both RSF and DeepSurv models significantly outperform the multivariate Cox model (precision:0.84, recall: 0.87, AUC: 0.86, C index: 0.87 for the DeepSurv model, while precision:0.89, recall: 0.87, AUC: 0.85, C index: 0.86 for RSF model) based on the same validation inputs of the risk predictors (P for trend <0.001). In addition, directly using the Cox model demonstrated better performance than the multivariate Cox model (precision:0.78, recall: 0.77, AUC: 0.76, C index: 0.75). The advantages of machine/deep learning approaches over the Cox model arise from the fact of their strength to describe survival data with both linear and nonlinear effects from covariates.

However, it should be noted that in comparison to DeepSurv, RSF allows influential predictors to be identified more easily by generating an 'importance ranking' of the variables with standard bootstrap theory. This enables the investigation of the predictive strength of associated risk predictors for clinicians to estimate the mortality probability by just referring to the most important variables.

Table 18. Survival prediction performance comparison between Cox, random survival forest and DeepSurv model with five-fold cross-validation approach

	Precision	Recall	AUC	C-index
Multivariate Cox	0.77	0.75	0.72	0.73
Cox score	0.78	0.77	0.76	0.75
DeepSurv	0.84	0.87	0.86	0.87
RSF	0.89	0.87	0.85	0.86

4.4. Discussion and limitations

4.4.1. Discussion

In this study, we developed a machine learning-driven predictive risk model T2DM using a multi-parametric approach with data from different domains. Our novel findings report that 1) measures of variability of fasting glucose and HbA1c show similar predictive power for all-cause mortality, regardless of whether adjustments were made for initial values or mean values across follow-up; 2) a multi-parametric predictive risk model incorporating variables from different domains, including baseline demographics, comorbidities and laboratory tests, measures of variability of HbA1c and fasting blood glucose predicted all-cause mortality accurately and 3) machine learning-driven algorithms further improved the accuracy of the predictive models.

Numerous factors have been associated with premature mortality in patients with T2DM. Prior epidemiological studies have identified key risk factors including age, comorbidities, healthcare utilization patterns and laboratory findings (177, 178). In our study, we also identified similar predictors that included advanced age, male gender, high neutrophil and low lymphocyte count, increased levels of urea, creatinine and potassium, as well as reduced levels of HDL-C, LDL-C, triglycerides, total cholesterol and sodium. Moreover, J-shaped relationships between LDL-C, HDL-C and total cholesterol were found in our cohort. These findings are in keeping with U-shaped relationships between cholesterol and all-cause mortality (179) and LDL-C (180) in the general Korean populations. Similar relationships were found for HDL-C, where extremely high LDL-C levels were paradoxically associated with higher mortality (181). The association between all-cause mortality and elevated creatinine, urea and potassium, which are classic features of renal failure, is supported by evidence suggesting that the Asian population has a higher risk of developing diabetic nephropathy compared with Caucasians (182). It is widely accepted that current predictive models that have largely been developed using Western cohorts only provide moderate levels of accuracy and at times do not lend themselves relevant to disease management protocols that vary by country. The development of country/ territory-specific risk prediction models allows for local population-based confounders and clinician management approaches to be incorporated into these models thus providing a more accurate risk prediction for the local population.

Diabetes is characterized by the presence of systemic chronic inflammation, which is accompanied by increased oxidative stress. To quantify the degree of inflammation, the NLR has been used as a surrogate measure, as it reflects the balance between pro- and anti-inflammatory pathway activation. In our cohort, we found that raised NLR was associated with all-cause mortality risk. We extend previous findings of our group and other groups that increased NLR has been associated with insulin resistance in patients newly diagnosed with

T2DM (183), the progression of diabetic nephropathy (184), and complications in diabetes. Consequently, the increased oxidative stress environment in diabetes can induce adverse remodelling of the heart, which in turn increases the risk of HF, arrhythmias, and cardiovascular mortality (82, 185).

Glycaemic variability refers to fluctuations in glucose levels and can be measured as a daily variation or variations between different clinical visits (8). Similarly, variability in HbA1c levels has been quantified. Both measures have been associated with a higher risk of complications and mortality in patients with diabetes mellitus in both randomised controlled trials and real-world settings (61, 127, 128, 186). Several methods can be used to calculate variability, such as SD, CV and score based on the frequency exceeding a fixed percentage change in the absolute values. Prior studies have demonstrated the importance of such measures of variability in the prediction of adverse outcomes (166, 167), but a systematic and direct comparison of different methodologies has not been made concerning their predictive performance. In our study, eight different measures of variability for HbA1c and FBG were compared, all of which showed significant predictive values. Our findings illustrate that temporal variability in these laboratory tests is important, regardless of the methodology employed for its calculation. In our study, we also found that mean FBG did not predict mortality. Instead, all the different measures of its variability were predictive, suggesting that it is intermittent poor glucose control rather than chronic hypoglycaemia is more closely associated with all-cause mortality.

A standard survival model such as the Cox proportional hazards model is a semiparametric analysis model to calculate the effects of observed patient covariates on the mortality risk outcome. The Cox model assumes the effect of each covariate is proportional. However, in many practical applications, the assumption is not true and risks losing decision

information among the observed patient's covariates. Furthermore, it cannot account for the presence of U-shaped relationships as only a single hazard ratio is derived for each covariate. Therefore, numerous nonlinear survival models were developed to better fit survival data with nonlinear log-risk functions (e.g., time-encoded methods (187)) or learning the nonlinear relationship directly using machine learning and deep learning techniques (e.g., feed-forward neural network risk-predicting methods (188)). RSF model (108) which is constructed by an ensemble of binary decision trees has been identified as an alternative approach to the Cox proportional hazards model in analysing time-to-event survival data when the linear proportional hazards assumption is violated. DeepSurv (189) whose multi-layer perceptron architecture is deeper than Faraggi-Simon's feed-forward model and minimizes the negative log Cox partial likelihood with a risk not necessarily linear, is capable of efficiently learning complex non-linear relationships between patient's covariates and mortality outcome. For model selection among the traditional Cox model, the Cox-based score model, RSF, and DeepSurv in risk prediction tasks, there exists a trade-off: (1) traditional Cox models (as well as Cox-based score models) provide good model interpretation ability but less accurate predictions since they sacrificed the consideration of nonlinear inter-dependent patterns among the variables; (2) machine learning or deep learning based models significantly improves prediction performance especially when the size of instance cohort is quite large (n >1000) but some of them (e.g., DeepSurv) may not provide good interpretations about the resulting predictions.

Recurrent neural network is another type of machine learning model. However, in recurrent neural network, feedback loops are implemented in the sequential prediction. The variables implemented in the model are not time-series data, therefore in the absence of specific time-of-occurrence for the clinical variables, the model devised would be quite arbitrary. In a feedforward neural network, such as the DeepSurv model that is used in the present study, the

variables can be viewed to be stand-alone events, therefore are easier to interpret and a better fit to the present analysis. Prediction accuracy and model interpretability are the two most important consideration for risk prediction model selection for clinical or medical use. This study demonstrates the superiority of adopting RSF model for the risk prediction due to both its highest prediction accuracy and good model interpretability.

The findings of this study illustrate that the machine/ deep learning model can better capture the highly complex and nonlinear relationships between prognostic variables and an individual patient's risk of mortality without prior variable selection or domain knowledge, compared with the traditional Cox analysis model. The application of machine/deep learning to survival analysis performs much better than the standard Cox model in predicting the mortality risk of diabetes mellitus patients. Additionally, machine/deep survival learning models will enable clinicians to provide personalised survival estimations based on the computed probability of mortality risk. In practice, medical researchers can use machine/deep survival learning models to improve overall survival prediction performance based on prognostic characteristics of diabetes mellitus patients and subsequently inform early efficient treatment options and even reduce mortality risk.

4.4.2. Strengths and limitations

The following strengths of our study should be noted. Firstly, this was a territory-wide study with large patient numbers with complete and long follow-up of mortality over 10 years, owing to the linkage of the electronic health records to the death registry. Secondly, the availability of different data types including prior comorbidities, laboratory test results that included longitudinal data and drug details meant that we were able to build a comprehensive

risk model for accurate prediction. Thirdly, the application of the latest machine learning techniques was able to further improve the risk predictions of the models.

However, there are some limitations which should be noted. Firstly, this was a retrospective study and therefore carries the potential bias, such as information bias, that is found in all studies of this type. Secondly, as with all studies using administrative databases, under-coding is a possibility. This was nevertheless mitigated by our definition of diabetes to include not only patients with the appropriate ICD coding but also those who were on any diabetic medication or met the criteria of diabetes by either HbA1c or fasting glucose results. Further research is needed to explore the potential for the present findings to be extrapolated onto type 1 diabetic patients. Thirdly, although the deep neural network survival learning approach demonstrates significant potential in providing much higher accurate predictions, the model's weak interpretability becomes the main obstacle to its real application in clinical practices. Investigations of developing interpretable deep survival learning models that provide highly accurate predictions with supportive explanations for diabetes mellitus patients become our next research concentration.

4.5. Conclusion

A multi-parametric model incorporating variables from different domains predicted all-cause mortality accurately in T2DM and a machine/ deep learning-driven approach provided further improvements for risk prediction.

Chapter 5. Risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

5.1. Introduction

SCD is a leading cause of death worldwide, accounting for approximately 25% of deaths of cardiovascular origin. (190) By contrast, AF, the most common sustained arrhythmia among adults, is increasingly prevalent around the world, particularly in developed countries. (191, 192) Significant increases in morbidity and mortality amongst large-scale epidemiological studies have been demonstrated amongst AF patients, including an increase in SCD risk. (193, 194, 195, 196) Furthermore, T2DM increases the risk of SCD, as demonstrated by a recent meta-analysis of population-based prospective studies. (197)

With the global shift towards a more personalised approach in the management of diabetes, there is an increasing interest in exploring the application of new parameters, such as HbA1c and lipid variability, to better monitor disease progression and evaluate the prognosis. Although the exact mechanisms remain unclear, increased long-term glycaemic and lipid variability is hypothesized to lead to endothelial dysfunction via an increase in oxidative stress. (198, 199, 200) Since haemoglobin has an average lifespan of 100 days, HbA1c can reflect glycaemic control in recent months. Therefore, HbA1c variability is not affected by short-term glycaemic changes due to diet and medication changes, thus it is a better representation of long-term glycaemic variation. However, existing studies have focused on risk prediction of all-cause mortality and general cardiovascular adverse events, with a limited number of studies exploring specifically arrhythmic risks amongst diabetics. (57, 77, 99) Moreover, those type 2 diabetics who are partially or fully dependent on insulin are more likely to have severe disease and may be at higher risk of arrhythmias.

The present study aimed to assess the predictive value of HbA1c and lipid variability towards SCD attempt, as well as incident AF in type 2 diabetic patients receiving insulin therapy.

5.2. Methods

5.2.1. Ethics approval and method overview

This study is a retrospective territory-wide observational study approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee (Application reference: 2018.462, 2018.643, 2019.361 [approval date: 15th August 2019]). The study cohort contains type 2 diabetic patients with insulin prescribed from any hospitals and outpatient clinics under the Hong Kong Hospital Authority from January 1st, 2009 to December 31st, 2009. The study cohort from 10 years ago was selected to ensure there si adequate follow-up given the retrospective nature of the study. Clinical and biochemical data of eligible patients were obtained through CDARS. CDARS has been used by our team and other teams to conduct population-based studies on different cardiovascular diseases (68, 105, 106, 170), including diabetes mellitus (112, 201, 202, 203), in the past.

5.2.2. Patient data

Clinical and biochemical data of the present cohort were extracted from CDARS. The outcomes of the present study are the occurrence of SCD and AF from January 1st, 2009 to December 31st, 2019. SCD attempt is defined as episodes of ventricular tachycardia (VT), ventricular fibrillation (VF), or non-specific cardiac arrest, which were diagnosed under clinical judgement with electrocardiographic or biochemical findings and subsequently coded into hospital records. VT/ VF on electrophysiological study is not included. Hypoglycaemia-

induced cases of SCD attempt or AF were defined as cases with dextrose infusion during the admission episode or had blood glucose measured ≤ 3.9 mg/mmol. Demographic details, including age and sex, were extracted. Patients were categorized into four groups based on their age: below age 55, between ages 55-64, between ages 65-74, above and include age 75. The number of baseline acute hospitalization episodes between January 1st, 2004 to December 31st, 2008 was also obtained. Furthermore, the average daily dose of different classes of cardiovascular medications and anti-diabetic agents was calculated by averaging the multiple between the daily dose frequency and drug dose by all patients with prescriptions of the specific drug class. Eight classes of anti-diabetic agents were examined: insulin, sulphonylurea, biguanide, alpha-glucosidase inhibitor, thiazolidinedione, DPP4I, GLP-1A, and meglitinide. Data on five classes of cardiovascular drugs were obtained: ACEI/ ARB, beta-adrenergic inhibitor, CCB, diuretics, and lipid-lowering agents.

In terms of patient comorbidities, the number of non-diabetic comorbidities and diabetes-related complications between January 1st, 1999 to December 31st, 2008 were obtained. The specific diabetic-related complications recorded include 1) amyotrophy, 2) arthropathy, 3) HHS/ DKA, 4) hypoglycaemia, 5) neuropathy, 6) retinopathy/ maculopathy, 7) PVD/ peripheral angiopathy, 8) nephropathy. Patient's past medical history of the following conditions that initiated between January 1st, 1999 to December 31st, 2008 were also extracted: 1) CKD, 2) COPD, 3) CLD, 4) HF, 5) IHD, 6) hypertension, 7) AMI, 8) stroke. ICD-9 codes were used to extract the study outcomes and pre-existing comorbidities, whilst the diabetic complications were extracted using the ICD-9-based Hospital Authority Master Disease Code Table (HAMDCT) for greater specificity.

Baseline data of urinalysis, renal and liver function tests, complete blood count, and other blood tests within the year 2008 were extracted. Urinalysis results include: 1) urine albumin/ creatinine ratio, 2) creatinine clearance, 3) 24-hour total urine protein and albumin,

4) spot urine protein, albumin, and glucose. Indices from complete blood count include: 1) the absolute number of haemoglobin, basophil, eosinophil, platelet, red and white blood cells, 2) MCV, 3) MCHC, 4) MCH, 5) haematocrit. The presence of anaemia, defined by sex-based thresholds of below 13g/dL for males and 12g/dL for females, was obtained. Blood test results extracted include 1) serum creatinine, 2) serum sodium and potassium, 3) serum urea and urate, 4) total serum protein and albumin, 5) total serum bilirubin, ALT and ALP, 6) FBG and random blood glucose. The presence of hypoglycaemia at baseline and the frequency of hypoglycaemic episodes were extracted. Hypoglycaemia was defined by fasting or random blood glucose below 3.9 mg/mmol.

The following blood results between January 1st, 2004 to December 31st, 2008 were extracted for the evaluation of their mean, variability and baseline value: 1) total cholesterol, 2) HDL-C, 3) LDL-C, 4) total triglyceride and 5) HbA1c. LDL-C results included findings from both direct and calculated measurements. Variability analysis of a biochemical index was only executed on patients with at least three measurements.

5.2.3. Statistical analysis

Statistical analysis was performed using R Studio, and statistical significance was defined as P-value < 0.05. Kaplan-Meier survival curve was used to portray the difference in actual or aborted SCD and AF survival between patients of different age groups, with the statistical significance of the intergroup difference evaluated using the log-rank test. Temporal variability of HbA1c and lipid indices were evaluated using calculated parameters of SD and CV. CV was measured by multiplying 100-fold the value calculated by the ratio SD and mean. SD and CV were used to measure variability since they were less affected by outliers. Whilst SD is independent of the mean, CV is independent of the scale thus more sensitive to small changes to the mean. To identify predictors for shorter time to aborted or actual SCD and AF

occurrence, univariate Cox regression was first applied to clinical and biochemical parameters. Patients with missing data were excluded from the analysis. Furthermore, due to their limited prescription towards the study cohort, GLP-1A and meglitinide were not included. Subsequently, parameters with P-value < 0.10 were included in the multivariate Cox regression model. Only patients with no missing data for the selected parameters, and at least three measurements for the selected variability predictors, were included in the multivariate models. No data imputation was performed.

The inter-relations between HbA1c and lipid variability with intermittent hypoglycaemia were evaluated using logistic and Poisson regression. Logistic regression was also used to examine the relationship between baseline hypoglycaemia frequency and 1) occurrence of SCD attempt/ AF or 2) aborted or actual SCD/ AF episodes that were associated with hypoglycaemia. ORs were reported from logistic and Poisson regression, whereas HRs were reported from Cox regression, along with the 95% CIs.

5.3 Results

5.3.1. Baseline characteristics

The present study included 23329 patients (mean age= 64.3 ± 13.8, male= 50.8%, average mean HbA1c= 8.6 ± 1.3%, all-cause mortality=50.5%). The cohort was divided into four age groups: <55 (n= 5511), 55-64 (n= 5745), 65-74 (n=6032), and>75 (n=6041), with a significant intergroup difference in survival for both aborted or actual SCD ($p < 0.0001$) and incident AF ($p < 0.0001$; **Figure 11 and 12**). Gender differences were not explored in the present study. The baseline clinical characteristics of the cohort are summarized in **Table 19**. Patients had an average of 8.67 ± 7.81 distinct non-diabetic comorbidities and 0.66 ± 0.90 diabetic complications. On average, patients had a median of 3 (interquartile range= 5) episodes of acute hospital admissions between the years 2004-2008. The three commonest diabetic complications were retinopathy/ maculopathy (16.5%), nephropathy (14.7%) and hypoglycaemia (11.9%), as shown in **Table 19**. Other pre-existing comorbidities were hypertension (36.6%), IHD (16.7%), stroke (12.1%), HF (10.1%), CKD (9.0%), CLD (5.7%), AMI (5.3%) and COPD (3.4%). 43.8% of patients (n=10223, mean daily dose= 95.0±300mg/day) were on lipid-lowering agents over the course of follow up.

Figure 12. Kaplan-Meier survival curves of different age groups for sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

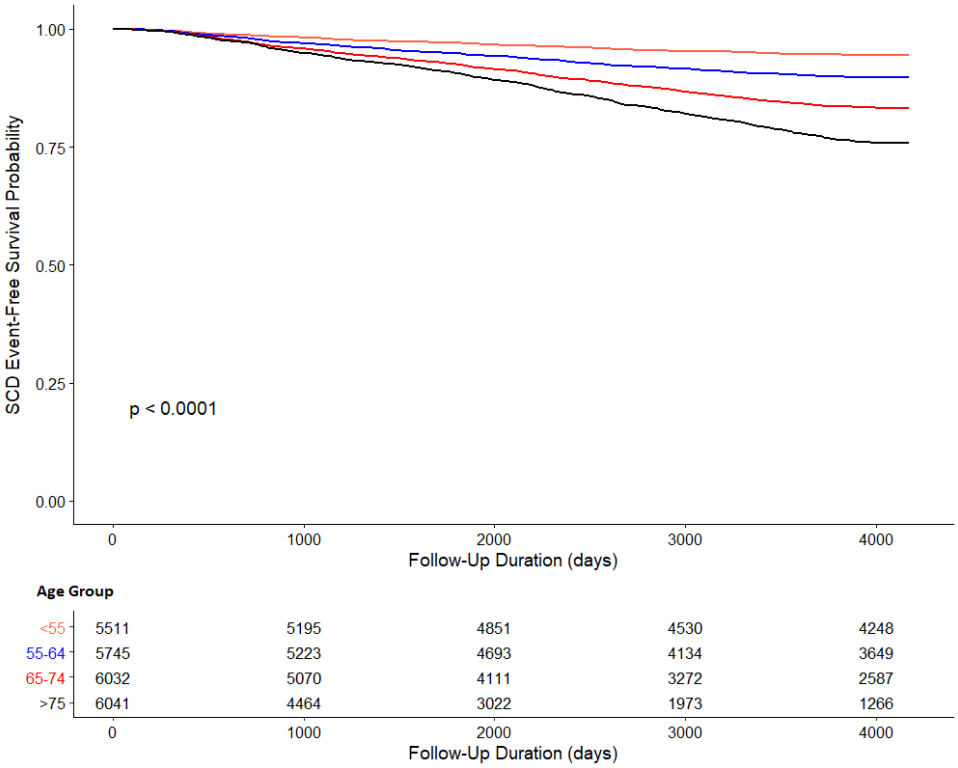


Figure 13. Kaplan-Meier survival curves of different age groups for atrial fibrillation in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

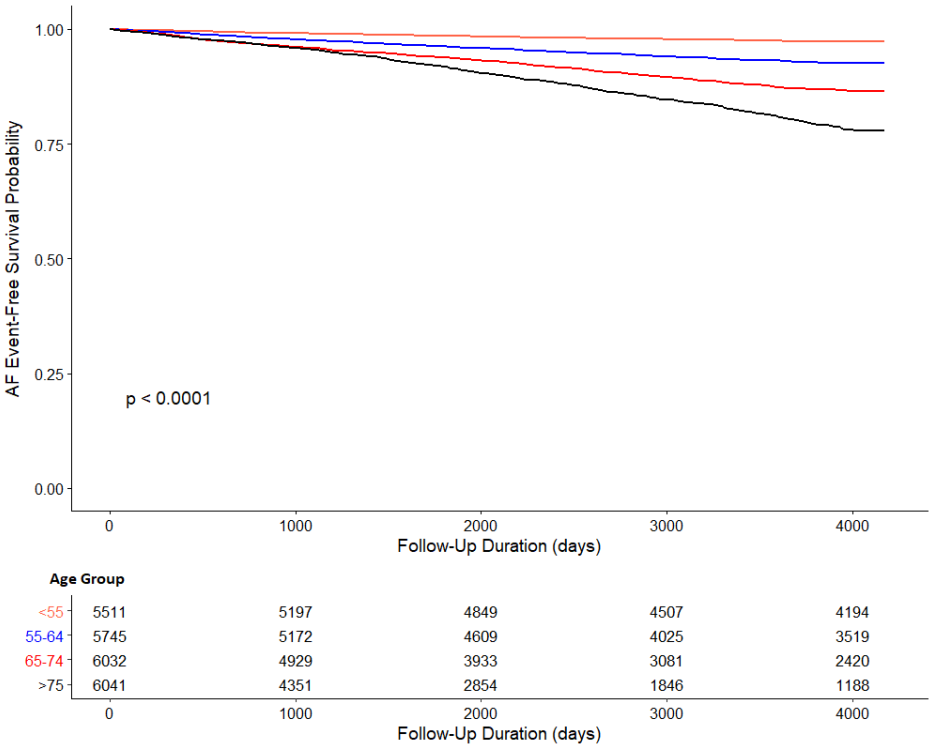


Table 19. Baseline clinical characteristics of risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

	n	%
Male	11842	50.8
Comorbidities		
Chronic Kidney Disease	2099	9.00
Chronic Obstructive Pulmonary Disease	783	3.36
Heart Failure	2355	10.1
Ischemic Heart Disease	3904	16.7
Hypertension	8538	36.6
Chronic Liver Disease	1335	5.72
Acute Myocardial Infarction	1224	5.25
Stroke	2826	12.1
Pre-existing Diabetic Complications		
Amyotrophy	70	0.300
Arthropathy	20	0.086
Hyperosmotic Hyperglycaemia State/ Diabetic Ketoacidosis	1267	5.43
Hypoglycaemia	2787	11.9
Neuropathy	1339	5.74
Retinopathy/ Maculopathy	3848	16.5
Peripheral Vascular Disease/ Angiopathy	544	2.33
Nephropathy	3420	14.7

The baseline biochemical characteristics of the present cohort are presented in **Table 20**. The average frequency of baseline hypoglycaemic episodes, between January 1st to December 31st 2008, was 0.50 ± 1.31 episodes per year. The baseline hypoglycaemia frequency significantly correlated with all HbA1c and lipid variability indices ($p < 0.0001$). Logistic regression revealed that baseline hypoglycaemia frequency was a significant predictor of aborted or actual SCD (OR= 1.09, 95% CI= [1.06, 1.12], $p < 0.0001$) and incident AF (OR= 1.05, 95% CI= [1.01, 1.08], p= 0.007). In total, 2512 and 1846 patients experienced incident SCD and AF respectively throughout follow-up.

Table 20. Baseline biochemical characteristics of risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

	Mean	Standard Deviation
Urinalysis		
Albumin/Creatinine Ratio (mg/mmol)	40.3	125
Creatinine Clearance (ml/min)	52.9	35.2
Spot Protein (g/d)	1.19	1.96
Spot Albumin (mg/L)	179	555
Spot Glucose (mmol/L)	12.5	6.54
24-hours Total Protein (g/d)	1.20	1.99
24-hours Total Albumin (mg/d)	279	677
Baseline Blood Test		
Fasting Glucose (mmol/L)	8.90	3.66
Random Glucose (mmol/L)	12.2	7.36
HbA1c (%)	8.56	1.93
Total Cholesterol (mmol/L)	4.74	1.12
High Density Lipoprotein (HDL) Cholesterol (mmol/L)	1.22	0.387
Calculated Low Density Lipoprotein (LDL) Cholesterol (mmol/L)	2.74	0.930
Direct LDL Cholesterol (mmol/L)	2.81	0.924
Triglyceride (mmol/L)	1.83	1.74
Thyroid-Stimulating Globulin (TSH) (mIU/L)	2.28	4.13
Free Thyroxine (fT4) (pmol/L)	14.7	37.3
Renal Function Test		
Creatinine (umol/L)	146	160
Sodium (mmol/L)	139	3.33
Potassium (mmol/L)	4.31	0.507
Urate (umol/L)	0.412	0.128
Urea (mmol/L)	8.98	6.11
Liver Function Test		
Albumin (g/L)	39.1	5.53
Alanine Aminotransferase (ALT) (U/L)	24.3	21.0
Alkaline Phosphatase (ALP) (U/L)	84.5	45.8
Total Bilirubin (umol/L)	11.2	9.07
Total Protein (g/L)	74.4	7.14
Complete Blood Count		
Haemoglobin (g/dL)	12.5	1.99
Mean Corpuscular Haemoglobin (MCH) (pg)	29.7	2.95
Mean Corpuscular Haemoglobin Concentration (MCHC) (g/dL)	34.0	0.953
Mean Corpuscular Volume (MCV) (fL)	87.2	7.42
Hematocrit (L/L)	0.376	0.559
Basophil ($x10^9$/L)	0.029	0.042
Eosinophil ($x10^9$/L)	0.225	0.236
Lymphocyte ($x10^9$/L)	1.87	0.866
Monocyte ($x10^9$/L)	0.539	0.266
Neutrophil ($x10^9$/L)	5.46	2.72

Platelet ($x10^9$/L)	255	83.4
Red Blood Cell ($x10^{12}$/L)	4.25	0.740
White Blood Cell ($x10^9$/L)	8.09	2.86

5.3.2. Predictors of sudden cardiac death and atrial fibrillation

In the present cohort, 10.3% of patients suffered from at least one SCD attempt, and 7.7% suffered from AF. Amongst these cases, 25 and 10 cases of patients experiencing aborted or actual SCD and AF were associated with hypoglycaemia on admission, respectively. **Tables 21** and **22** present the univariate Cox regression of predictors for aborted or actual SCD and AF respectively. **Table 23** presents the univariate Cox regression for VT alone, with HbA1c and cholesterol variability found to be predictive ($P < 0.05$).

Table 21. Univariate predictors for sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

Predictor	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.046	[1.04, 1.05]	< **0.0001**
Categorized Age	1.61	[1.55, 1.67]	<**0.0001**
Male	1.27	[1.17, 1.37]	<**0.0001**
Frequency of Baseline Acute Admissions	1.005	[1.00, 1.01]	<**0.0001**
Number of Concomitant DM Complications	1.45	[1.39, 1.51]	<**0.0001**
Number of Distinct Non-DM Comorbidities	1.053	[1.05, 1.06]	<**0.0001**
Baseline Haemoglobin Count	1.01	[0.989, 1.03]	0.367
Baseline Anaemia	0.930	[0.854, 1.01]	0.094
Hypoglycaemia Frequency	1.11	[1.08, 1.13]	<**0.0001**
HbA1c (n=20874)			
Baseline	1.00	[0.981, 1.02]	0.870
Mean	0.997	[0.965, 1.03]	0.834
Standard Deviation	1.10	[1.06, 1.15]	<**0.0001**
Coefficient of Variation	1.01	[1.006, 1.02]	<**0.0001**
Total Cholesterol (n=18926)			
Baseline	0.973	[0.938, 1.01]	0.154
Mean	1.02	[0.973, 1.08]	0.362
Standard Deviation	1.33	[1.25, 1.41]	<**0.0001**
Coefficient of Variation	1.03	[1.02, 1.03]	<**0.0001**
HDL Cholesterol (n=17930)			

Baseline	1.00	[0.902, 1.12]	0.943
Mean	0.471	[0.406, 0.547]	**<0.0001**
Standard Deviation	2.54	[1.66, 3.90]	**<0.0001**
Coefficient of Variation	1.024	[1.02, 1.03]	**<0.0001**
LDL Cholesterol (n=17485)			
Baseline	1.01	[0.962, 1.06]	0.746
Mean	0.990	[0.929, 1.06]	0.749
Standard Deviation	1.66	[1.48, 1.85]	**<0.0001**
Coefficient of Variation	1.015	[1.01, 1.02]	**<0.0001**
Triglyceride (n=18889)			
Baseline	1.01	[0.991, 1.03]	0.251
Mean	1.08	[1.06, 1.10]	**<0.0001**
Standard Deviation	1.05	[1.02, 1.08]	**0.001**
Coefficient of Variation	1.00 (1.004)	[1.00, 1.01]	**0.001**
Anti-Diabetic Agent			
Sulphonylurea	1.18	[1.09, 1.28]	**<0.0001**
Biguanide	0.464	[0.428, 0.503]	**<0.0001**
DPP4 Inhibitor	0.587	[0.293, 1.18]	0.132
Thiazolidinedione	0.663	[0.519, 0.847]	**0.001**
Alpha-Glucosidase Inhibitor	1.10	[0.884, 1.37]	0.393

Table 22. Univariate predictors for atrial fibrillation in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

Predictor	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.059	[1.05, 1.06]	**<0.0001**
Categorized Age	1.86	[1.77, 1.95]	**<0.0001**
Male	0.903	[0.823, 0.991]	**0.031**
Frequency of Baseline Acute Admissions	1.004	[1.00, 1.01]	**<0.0001**
Number of Concomitant DM Complications	1.20	[1.13, 1.27]	**<0.0001**
Number of Distinct Non-DM Comorbidities	1.034	[1.03, 1.04]	**<0.0001**
Baseline Haemoglobin Count	0.981	[0.957, 1.01]	0.131
Baseline Anaemia	1.02	[0.925, 1.13]	0.674
Hypoglycaemia Frequency	1.08	[1.05, 1.11]	**<0.0001**
HbA1c (n=20874)			
Baseline	1.00	[0.978, 1.03]	0.874
Mean	0.937	[0.902, 0.973]	**0.001**
Standard Deviation	1.09	[1.04, 1.15]	**0.001**
Coefficient of Variation	1.01	[1.005, 1.02]	**<0.0001**
Total Cholesterol (n=18926)			
Baseline	1.02	[0.982, 1.07]	0.264
Mean	0.872	[0.820, 0.927]	**<0.0001**
Standard Deviation	1.18	[1.08, 1.28]	**<0.0001**

Coefficient of Variation	1.016	[1.01, 1.02]	<**0.0001**
HDL Cholesterol (n=17930)			
Baseline	0.965	[0.852, 1.09]	0.573
Mean	0.666	[0.566, 0.783]	<**0.0001**
Standard Deviation	1.82	[1.09, 3.05]	**0.022**
Coefficient of Variation	1.024	[1.02, 1.03]	<**0.0001**
LDL Cholesterol (n=17485)			
Baseline	0.975	[0.924, 1.03]	0.343
Mean	0.792	[0.734, 0.854]	<**0.0001**
Standard Deviation	1.31	[1.14, 1.51]	<**0.001**
Coefficient of Variation	1.015	[1.01, 1.02]	<**0.0001**
Triglyceride (n=18889)			
Baseline	0.975	[0.944, 1.01]	0.123
Mean	1.04	[1.01, 1.07]	**0.018**
Standard Deviation	0.996	[0.954, 1.04]	0.850
Coefficient of Variation	0.998	[0.995, 1.00]	0.218
Anti-Diabetic Agent			
Sulphonylurea	1.10	[1.00, 1.21]	**0.045**
Biguanide	0.682	[0.621, 0.748]	<**0.0001**
DPP4 Inhibitor	0.585	[0.262, 1.30]	0.190
Thiazolidinedione	0.816	[0.632, 1.05]	0.119
Alpha-Glucosidase Inhibitor	0.972	[0.743, 1.27]	0.836

Table 23. Univariate predictors for ventricular tachycardia in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

Predictor	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.03	[1.02, 1.05]	<**0.0001**
Categorized Age	1.43	[1.26, 1.63]	<**0.0001**
Male	1.93	[1.44, 2.59]	<**0.0001**
Frequency of Baseline Acute Admissions	1.01	[1.00, 1.01]	<**0.0001**
Number of Concomitant DM Complications	1.26	[1.08, 1.46]	**0.003**
Number of Distinct Non-DM Comorbidities	1.06	[1.04, 1.07]	<**0.0001**
Baseline Haemoglobin Count	0.980	[0.912, 1.05]	0.576
Baseline Anaemia	0.992	[0.743, 1.32]	0.956
Hypoglycaemia Frequency	0.986	[0.878, 1.11]	0.806
HbA1c (n=20874)			
Baseline	1.02	[0.908, 1.06]	0.599
Mean	0.941	[0.842, 1.05]	0.285
Standard Deviation	1.13	[0.973, 1.31]	0.111
Coefficient of Variation	1.02	[1.00, 1.03]	**0.031**
Total Cholesterol (n=18926)			
Baseline	1.00	[0.877, 1.13]	0.954
Mean	1.02	[0.822, 1.16]	0.789

Standard Deviation	1.29	[1.05, 1.59]	**0.018**
Coefficient of Variation	1.02	[1.01, 1.04]	**0.008**
HDL Cholesterol (n=17930)			
Baseline	1.03	[0.708, 1.48]	0.897
Mean	3.46	[0.168, 0.497]	**<0.0001**
Standard Deviation	1.52	[0.335, 6.87]	0.589
Coefficient of Variation	1.03	[1.01, 1.04]	**0.002**
LDL Cholesterol (n=17485)			
Baseline	0.954	[0.812, 1.12]	0.566
Mean	1.01	[0.819, 1.25]	0.923
Standard Deviation	1.62	[1.12, 2.35]	**0.010**
Coefficient of Variation	1.01	[1.00, 1.02]	**0.024**
Triglyceride (n=18889)			
Baseline	1.02	[0.963, 1.09]	0.441
Mean	1.09	[1.03, 1.16]	**0.003**
Standard Deviation	1.06	[0.969, 1.15]	0.211
Coefficient of Variation	1.01	[0.999, 1.01]	0.088
Anti-Diabetic Agent			
Sulphonylurea	1.26	[0.961, 1.66]	0.094
Biguanide	0.494	[0.377, 0.650]	**<0.0001**
DPP4 Inhibitor	/	/	/
Thiazolidinedione	0.831	[0.391, 1.77]	0.629
Alpha-Glucosidase Inhibitor	0.757	[0.312, 1.84]	0.539

Table 24 summarizes the results of the multivariate analysis for aborted or actual SCD (n=15316), where the following predictors were independent predictors: 1) demographics: age (HR= 1.04, 95% CI= [1.02, 1.05], p < 0.0001), male (HR= 1.27, 95% CI= [1.15, 1.41], p < 0.0001); 2) clinical: frequency of baseline acute admissions (HR= 1.003, 95% CI= [1.003, 1.003], p < 0.0001), number of concomitant diabetic complications (HR= 1.21, 95% CI= [1.15, 1.27], p < 0.0001), number of discrete non-diabetic comorbidities (HR= 1.02, 95% CI= [1.01, 1.02], p < 0.0001), hypoglycaemic frequency (HR= 1.03, 95% CI= [1.00, 1.07], p= 0.032); 3) glycaemic and lipid variability: HbA1c SD (HR= 1.74, 95% CI= [1.45, 2.09], p < 0.0001) and CV (HR= 0.953, 95% CI= [0.935, 0.972], p < 0.0001), total cholesterol CV (HR= 1.05, 95% CI= [1.02, 1.08], p= 0.002), LDL-C SD (HR= 1.73, 95% CI= [1.10, 2.74], p=0.018), triglyceride mean (HR= 1.29, 95% CI= [1.17, 1.42], p < 0.0001) and CV (HR= 1.01, 95% CI= [1.00, 1.01], p= 0.022); 4) antidiabetic agent: sulphonylurea (HR= 1.39, 95% CI= [1.25, 1.54],

p < 0.0001), biguanide (HR= 0.630, 95% CI= [0.563, 0.704], p < 0.0001); 5) cardiovascular medications: beta-blocker (HR= 1.24, 95% CI=[1.12, 1.38], p < 0.0001), diuretic (HR= 1.51, 95% CI= [1.35, 1.67], p < 0.0001). CV of HbA1c (HR= 0.953, 95% CI= [0.935, 0.972], p < 0.0001) and LDL-C (HR= 0.978, HR= [0.966, 0.991], p= 0.001), in addition to SD of total cholesterol (HR= 0.551, 95% CI= 0.275, 0.951), p=0.034) and triglyceride (HR= 0.772, 95% CI= [0.656, 0.909], p=0.002) were predictive of SCD on univariate analysis but became protective after multivariate analysis.

Table 24. Multivariate analysis showing predictors of sudden cardiac death (n=15316) in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

Predictor	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.04	[1.03, 1.06]	**<0.0001**
Categorized Age	0.928	[0.806, 1.07]	0.295
Male	1.31	[1.18, 1.45]	**<0.0001**
Frequency of Baseline Acute Admissions	1.002	[1.001, 1.003]	0.295
Number of Concomitant DM Complications	1.23	[1.17, 1.29]	**0.001**
Number of Distinct Non-DM Comorbidities	1.026	[1.02, 1.03]	**<0.0001**
Baseline Anaemia	0.919	[0.834, 1.01]	0.084
Hypoglycaemia Frequency	1.05	[1.02, 1.08]	**<0.0001**
HbA1c			
Standard Deviation	1.45	[1.21, 1.75]	**<0.0001**
Coefficient of Variation	0.970	[0.952, 0.989]	**0.002**
Total Cholesterol			
Standard Deviation	0.779	[0.421, 1.44]	0.427
Coefficient of Variation	1.04	[1.00, 1.07]	**0.031**
HDL Cholesterol			
Mean	0.780	[0.554, 1.10]	0.156
Standard Deviation	0.811	[0.115, 5.71]	0.834
Coefficient of Variation	1.01	[0.984, 1.03]	0.539
LDL Cholesterol			
Standard Deviation	1.42	[0.912, 2.21]	0.120
Coefficient of Variation	0.984	[0.972, 0.997]	**0.014**
Triglyceride			
Mean	1.36	[1.24, 1.49]	**<0.0001**
Standard Deviation	0.684	[0.579, 0.808]	**<0.0001**
Coefficient of Variation	1.007	[1.00, 1.01]	**0.012**
Anti-Diabetic Agent			

Sulphonylurea	1.39	[1.25, 1.54]	<**0.0001**
Biguanide	0.584	[0.523, 0.651]	<**0.0001**
Thiazolidinedione	0.870	[0.669, 1.13]	0.299

The findings of multivariate regression analysis for incident AF (n=13267) are presented in **Table 25**, where several significant predictors were identified: 1) demographics: age (HR= 1.04, 95% CI= [1.03, 1.06], p < 0.0001)2) clinical: frequency of baseline acute admissions (HR= 1.00, 95% CI=[1.00, 1.00], p < 0.0001), hypoglycaemic frequency (HR= 1.05, 95% CI=[1.02, 1.09], p=0.005); 3) antidiabetic agent: sulphonylurea (HR= 1.12, 95% CI= [1.00, 1.26], p= 0.043), biguanide (HR= 0.843, 95% CI= [0.746, 0.954], p= 0.007); 4) cardiovascular medication: ACEI/ARB (HR= 1.21, 95% CI= 1.06, 1.37), p=0.004), beta-blocker (HR= 1.56, 95% CI=[1.39, 1.75], p < 0.0001), CCB (HR= 1.35, 95% CI= [1.20, 1.51], p < 0.0001), diuretic (HR= 1.50, 95% CI= [1.34, 1.69], p < 0.0001), lipid-lowering agents (HR= 1.16, 95% CI= [1.03, 1.32], p=0.018).

Table 25. Multivariate analysis showing predictors of atrial fibrillation (n=13267) in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study

Predictor	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.05	[1.03, 1.06]	<**0.0001**
Categorized Age	1.07	[0.893, 1.28]	0.466
Male	0.869	[0.762, 0.991]	**0.037**
Frequency of Baseline Acute Admissions	1.004	[1.00, 1.01]	<**0.001**
Number of Concomitant DM Complications	1.07	[0.991, 1.15]	0.084
Number of Distinct Non-DM Comorbidities	1.00	[0.994, 1.01]	0.459
Hypoglycaemia Frequency	1.08	[1.04, 1.12]	<**0.0001**
HbA1c			
Mean	1.02	[0.937, 1.11]	0.648
Standard Deviation	1.11	[0.743, 1.65]	0.620
Coefficient of Variation	1.01	[0.969, 1.04]	0.790
Total Cholesterol			
Mean	1.41	[0.937, 2.02]	0.065
Standard Deviation	0.635	[0.259, 1.56]	0.321
Coefficient of Variation	1.02	[0.978, 1.07]	0.316
HDL Cholesterol			

Mean	0.674	[0.409, 1.11]	0.121
Standard Deviation	0.340	[0.03, 4.47]	0.411
Coefficient of Variation	1.02	[0.986, 1.05]	0.268
LDL Cholesterol			
Mean	0.539	[0.370, 0.784]	**0.001**
Standard Deviation	2.95	[1.37, 6.39]	**0.006**
Coefficient of Variation	0.975	[0.955, 0.996]	**0.018**
Triglyceride			
Mean	0.947	[0.828, 1.08]	0.426
Anti-Diabetic Agent			
Sulphonylurea	1.17	[1.03, 1.34]	**0.015**
Biguanide	0.824	[0.717, 0.948]	**0.007**

5.4. Discussion and limitations

5.4.1. Discussion

The present study demonstrated the following major findings: 1) clinical and biochemical indices are predictive of arrhythmic occurrence amongst diabetics; 2) both the mean and variability of HbA1c and lipid indices can predict VT/ VF/ SCD in diabetic patients; 3) HbA1c variability is associated with hypoglycaemia frequency. To the best of our knowledge, the present study is the first to report an association between increased variability in HbA1c and lipid markers with increased risk for VT/ VF/ SCD amongst diabetic patients.

The prognostic values of HbA1c and lipid variability have been increasingly explored over the past decades. However, prior studies have mostly focused on the prediction of all-cause mortality or cardiovascular events, and studies on predictions for arrhythmogenesis were few. (57, 83, 99) Although the underlying pathophysiology remains incompletely elucidated, there are several possible contributing factors towards the increased arrhythmic risk amongst patients with high glycaemic and lipid variability. First, increased glycaemic variability is found to be associated with corrected QT interval (QTc) prolongation and increased QTc dispersion, which greatly elevates the risk of ventricular tachyarrhythmia. (204, 205) There is evidence suggesting that QTc prolongation may be triggered by spontaneous hypoglycaemia due to underlying coronary atherosclerosis or cardiac autonomic neuropathy (206, 207, 208, 209). Anti-diabetic agent use may also play a role in the prognostic value of glycaemic variability. Biguanide users are likely more stable or earlier in the disease course, and thus have a lower cardiovascular disease burden. Sulphonylurea use, which was predictive of SCD in the present study, is known to have an increased risk of hypoglycaemia. (210) In addition, it has been reported human ether-a-go-go-related gene (hERG) channel inhibitory effects of some sulphonylurea, which can lead to QT prolongation. (211) Amongst patients on insulin, who

have more labile glucose control, the spontaneous glycaemic fluctuations can induce the occurrence of arrhythmia. Unfortunately, continuous blood glucose monitoring was not available in the present study to demonstrate the association between spontaneous glycaemic changes and arrhythmic episodes.

Furthermore, structural remodelling may also be involved in the pathogenesis. HbA1c variability has been associated with left ventricular remodelling and atrial fibrosis, which could be arrhythmogenic. (81, 212, 213) In a recent nationwide observational study, an association between high lipid variability and increased risk of new-onset AF was reported, and statins protected against AF development via the reduction in adverse atrial remodelling. (123)Moreover, frequent intermittent hypoglycaemia can induce the release of reactive oxygen species (82), thereby leading to increased oxidative stress, chronic inflammation, and endothelial dysfunction. (80) Hypoglycaemia itself is arrhythmogenic and can reduce the myocardial tolerance to ischemia and reperfusion injuries. (214, 215)

Similar to glycaemic variability, the increase in oxidative stress with fluctuations in lipid levels due to atherogenic substance release from unstable plaque is hypothesized to underlie the increased arrhythmic risk. (99) Indeed, glycaemic fluctuations were found to increase the formation of atherosclerotic plaques and thinning of the fibrous cap, which suggests that intermittent hypoglycaemia may contribute to lipid variability as well. (216) It should be noted that whilst triglyceride and HDL-C variability are dependent on glycaemic control and other lifestyle factors, the use of statin plays a significant role in LDL-C variability. The significant interpersonal variability, as well as the varying effects between different types of statins on LDL-C variability, reflects the need for further research in the area. (217, 218)

The change in HbA1c and lipid variability from predictive of SCD, in univariate analysis, to being protective under multivariate analysis, can be attributed to several causes. The limitation of cohort size and multivariate analysis may have been selected for patients of

more advanced disease and undergone aggressive control. Given that a J-shaped association between adverse outcomes and both glycaemic and lipid indices have been described, patients with high variability that returned to the optimal glycaemic and lipid range would have had a better prognosis. (60, 113, 219) Indeed, the duration of exposure to an optimal glycaemic range is inversely associated with diabetic retinopathy progression, even after accounting for the effects of glycaemic variability. (220) Additionally, Ceriello *et al*. reported that patients with elevations in both HbA1c and HDL-C variability were at higher risk for diabetic nephropathy than those with high variability in only one variable, highlighting the interacting effects between variability markers. (95) Therefore, the protective value may be a result of inevitable selection bias, the protective effects of pharmacotherapy, and the interactions between different indices. (221, 222)

5.4.2. Strengths and limitations

There are four major strengths of the present study: 1) the independent, and interdependent predictive effects of clinical and biochemical indices towards SCD and AF were assessed by univariate and multivariate analysis; 2) the predictive values of both the value and variability of HbA1c and lipid indices were assessed to examine the effect of biochemical fluctuations on arrhythmic risk in diabetics; 3) the inter-relationship between intermittent hypoglycaemia, HbA1c and lipid variability and chronic inflammation was examined to elucidate the underlying pathogenic mechanism; 4) long follow-up durations permitted the capture of adverse outcomes over a long period.

However, several limitations should be recognized. Firstly, the cohort was limited to type 2 diabetic patients prescribed insulin, which can limit the generalizability of the findings. Given that insulin is only prescribed for diabetic patients in later stages, an advanced disease state can be inferred for the selected patients. Secondly, the observational nature of the present

study leads to inevitable errors from missing data, coding errors, and under-coding. A causational relationship cannot be established from the findings of the observational study, which can only demonstrate associations. Furthermore, unfortunately, ICD coding does not reflect the frequency of events, thus the frequency of SCD attempts was not evaluated. The coding also does not reflect whether the VT is sustained or associated with hemodynamic collapse. It is based on the assumption that coded VT is clinically significant as non-sustained VT would be clinically irrelevant and thus not coded into the database. It should be emphasized that a diagnosis of VT is not the same as SCD, hence the effect of long term glycaemic variability on the risk of SCD should be interpreted with caution. Novel therapies, such as GLPA (n=9) and SGLT-2 inhibitor (n=0), were not assessed since this retrospective study recruited patients in the year 2009, at which these agents were not yet developed. Finally, data on blood pressure, body mass index, echocardiogram, the severity of HF and lifestyle were absent, which can affect the patients' cardiovascular health. Gender differences were also not explored in the present study.

5.5. Conclusion

Poor glucose control and variability in lipid parameters in diabetic patients are associated with SCD. These observations suggest the need to re-evaluate the extent of glycaemic control required for outcome optimization. Further studies on the predictive value of variability in other glycaemic measures, such as FBG and random blood glucose, in addition to other methods of measuring variability, should be performed to further examine the predictiveness of glycaemic variability towards arrhythmias in diabetic patients.

Chapter 6. Predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death

6.1. Introduction

T2DM is an increasingly prevalent disease burden across the globe due to ageing and lifestyle westernization, with numbers projected to increase by up to 439 million by 2030. (223) Diabetes mellitus is burdensome to the healthcare system for its chronic course and a multitude of possibly debilitating and lethal complications across different organ systems. AMI and SCD are major cardiovascular adverse outcomes in patients with T2DM. (224, 225)

Given the potentially lethal and debilitating nature of such cardiovascular adverse outcomes, many risk scores have been developed in hopes of identifying high-risk patients for early intervention and close monitoring. For example, the UKPDS Risk Engine is a T2DM risk score based on the UKPDS for ischemic heart disease. (226) The Reynolds Risk Score was developed to assess female cardiovascular risk, and the China-PAR project was devised to target the Chinese population specifically. (227, 228) However, typically these risk scores involving HbA1c and lipid level predicted composite outcomes of major cardiovascular adverse outcomes or cardiovascular mortality, which did not account for the difference in pathogenesis and prognosis between acute coronary syndrome and lethal ventricular arrhythmias. Furthermore, recent studies reported that HbA1c and lipid levels, which were often accounted for in these risk scores, have J/ U-shaped relationships with adverse outcomes. (4, 60, 229, 230) Therefore, updated risk scores that incorporate these new findings for predictions of specific cardiovascular adverse outcomes were warranted for personalised management.

The present study evaluated the application of incorporating non-linear J/U-shaped relationships between both mean HbA1c and cholesterol levels into risk scores for predicting AMI and non-AMI related SCD respectively, amongst T2DM patients. A conditional inference

survival forests (CISF) model was used for time-to-event survival data analysis in predicting AMI and non-AMI SCD (231, 232).

6.2. Methods

6.2.1. Study design

The present study has been approved by The Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee. Patients fulfilling all of the following inclusion criteria were recruited: 1) above the age of 40; 2) had documented diagnosis of T2DM under the ICD-9 coding system, or prescribed anti-diabetic agents between January 1st, 2009 to December 31st, 2009 by any of the Hong Kong Hospital Authority-managed public hospitals or outpatient clinics; 3) without prior history of AMI and SCD episodes. The data was collected from CDARS. The system has been used for cohort studies by both the present research team and other teams in the past (67, 68, 128, 170).

6.2.2. Data extraction

The primary outcome of the present study, the time to the initial AMI and non-AMI related SCD episode, is defined as days from January 1st, 2009 to the date of initial AMI/ non-AMI related SCD or the end of the follow-up period (December 31st, 2019). An SCD episode is defined as an episode of sustained VT, VF, or non-specific cardiac arrest. This includes episodes that were aborted (sudden cardiac arrest), and episodes that resulted in death. SCD episodes with AMI within a week before or after the SCD episode were considered AMI-related and thus excluded. The number of AMI and non-AMI related SCD episodes during the follow-up period was extracted as well. Other clinical characteristics, including demographic details (age and sex), diabetes duration, pre-existing comorbidities, anti-diabetic agents, and

cardiovascular agents prescribed, and all-cause mortality were also extracted. The onset of diabetes is determined by fulfilment of the following criteria, whichever is the earliest: 1) earliest record of T2DM related ICD-9 codes; 2) earliest record of HbA1c > 6.5%; 3) earliest record of FBG > 7 mmol/L. The duration of diabetes is defined as the onset of diabetes until December 31st, 2009. Similarly, follow-up duration was defined as from January 1st, 2009 to December 31st, 2019 or the date of death.

The following pre-existing comorbidities were identified using ICD-9 codes: 1) renal, ophthalmological, and neurological diabetic comorbidities; 2) HF; 3) AF; 4) hypertension; 5) PVD; 6) ischemic stroke; 7) osteoporosis; 8) COPD; 9) IHD. The classes of anti-diabetic agents extracted were: 1) biguanide; 2) sulphonylurea; 3) insulin; 4) DPP4I; 5) GLP-1A; 6) meglitinide; 7) alpha-glucosidase inhibitor; 8) thiazolidinedione. Antihypertensives (ACEI/ARB, beta-adrenergic receptor blocker, CCB, diuretics) and lipid-lowering agents were also extracted.

Baseline laboratory data from complete blood count (lymphocyte, neutrophil count and haemoglobin level), liver function test (ALT, ALP, albumin and total protein), renal function test (creatinine, sodium, potassium, urea), lipid (HDL-C, LDL-C, total cholesterol, triglyceride) and glycaemic profile (FBG, HbA1c) between January 1st, 2008 to December 31st, 2008 were obtained. Baseline anaemia was defined as haemoglobin count < 13g/dL amongst males, and <12g/dL amongst females. Mean HbA1c and FBG from January 1st, 2004 to December 31st, 2008 were also calculated.

6.2.3. Statistical analysis

The annualized rate and mean event frequency were calculated for the primary outcomes. The annualized rate was calculated by dividing the total number of episodes across the cohort by the number of patient-year follow-ups. The mean event annual frequency was

calculated by averaging the individual mean number of episodes per year throughout follow-up amongst those who experienced the event. Univariate Cox regression was used to identify predictors for incident episodes of both AMI and non-AMI related SCD. Patients with AMI before non-AMI related SCD were excluded from the SCD analysis. HR, 95% CI, and P-value were reported for the Cox regression. Univariate predictors with P-value < 0.10 were entered into a multivariate model. Significant predictors were then selected into predictive scores. The multivariate Cox regression was then repeated with only the significant predictors to obtain the HR for adjustments for the score. For variables with HR between 0.67 to 1.5, a score of 1 was assigned, otherwise, a score of 2 was assigned.

To examine the potential incorporation of the J/U-shaped relationship reported between glycaemic/cholesterol profile and cardiovascular adverse events, the deciles of these parameters that were included in the score were obtained and used to derive the HR predicting for AMI and non-AMI related SCD respectively through univariate Cox regression. Then, the decile with the minimal HR, excluding the first and last decile, was selected as the reference decile and compared against the remaining deciles. Univariate Cox regression was then repeated, and the derived HR was plotted. Parameters that displayed a J/ U-shaped relationship with the selected outcome would have had the score adjusted for, with the minimum and maximum cut-offs derived deciles that had a statistically insignificant difference in HR with the reference decile. The cut-off for other continuous variables included in the score was derived by maximizing the sensitivity and specificity. To evaluate the scores, an ROC curve was then generated for the scores, and the AUC was calculated. Statistical significance was defined as p-value < 0.05. The statistical analysis was performed using RStudio software (Version: 1.1.456).

6.3. Results

6.3.1. Baseline characteristics

This study included 261308 patients (age= 66.0 ± 11.8 years old, male= 47.6%, follow-up duration= 3552 ± 1201 days, diabetes duration= 4.77 ± 2.29 years). The categorical and continuous baseline demographic, clinical, and laboratory features are presented in **Tables 26 and 27**, respectively. The mean HbA1c level was 7.67 ± 1.17%, with anaemia present in 14.3% of the cohort at baseline. On follow-up, 33.3% (n=86908) of the patients died. The five most prevalent comorbidities in decreasing order are hypertension (23.1%), IHD (7.7%), HF (3.5%), ischemic stroke (3.3%), and AF (2.8%). On average, patients had 0.44 ± 0.80 of the extracted comorbidities. In terms of drug use, most patients were on monotherapy or combination therapy of biguanide (69.0%), sulphonylurea (64.0%), and insulin (10.4%), on average on 1.45 ± 0.80 anti-diabetic agents. ACEI/ ARB (19.0%) was the most common class of antihypertensive prescribed, followed by CCB (17.4%) and beta-adrenergic receptor blocker (14.6%). Lipid-lowering agents were prescribed in 10.6% of the patients. On average, patients from the present cohort were on 1.58 ± 1.27 cardiovascular medications.

Table 26. Baseline characteristics for categorical variables in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death

Characteristics	Number (Percentage)		
	Total Cohort (n=261308)	Acute Myocardial Infarction (n=20419)	Sudden Cardiac Death (n=12282)
Male	124495 (47.6)	10221 (50.1)	6454 (52.5)
Mortality	86908 (33.3)	14374 (70.4)	12096 (98.5)
Acute Myocardial Infarction (AMI)	20419 (7.81)	-	-

Sudden Cardiac Death (SCD)	12282 (4.74)	-	-
Baseline Anaemia	37286 (14.3)	5048 (24.7)	3470 (28.3)
Anti-Diabetic Agent			
Biguanide	180232 (69.0)	13797 (67.6)	7776 (63.3)
Sulphonylurea	167174 (64.0)	14421 (70.6)	8684 (70.7)
Insulin	27269 (10.4)	3620 (17.7)	2115 (17.2)
Meglitinide	25 (0.010)	3 (0.015)	3 (0.024)
Dipeptidyl Peptidase-4 Inhibitor	316 (0.121)	22 (0.108)	10 (0.081)
Thiazolidinedione	3741 (1.43)	335 (1.64)	162 (1.32)
Glucagon-like Peptide-1 Agonist	15 (0.006)	0 (0)	0 (0)
Acarbose	3119 (1.19)	404 (1.98)	218 (1.77)
Cardiovascular Drugs			
Angiotensinogen converting enzyme inhibitor (ACEI)/ angiotensin receptor blocker (ARB)	49712 (19.0)	5769 (28.3)	3363 (27.4)
Beta-adrenergic receptor blocker	38144 (14.6)	4577 (22.4)	2524 (20.6)
Calcium Channel Blocker	45542 (17.4)	5604 (27.4)	3265 (26.6)
Diuretic	24204 (9.26)	3209 (15.7)	2079 (16.9)
Lipid-Lowering Agent	27828 (10.6)	3797 (18.6)	1932 (15.7)
Comorbidities			
Renal Diabetic Complication	3049 (1.17)	563 (2.76)	382 (3.11)
Peripheral Vascular Disease (PVD)	299 (0.114)	78 (0.382)	33 (0.269)
Ophthalmological Diabetic Complication	3255 (1.25)	627 (3.07)	376 (3.06)
Neurological Diabetic Complication	1066 (0.408)	191 (0.935)	116 (0.944)
Ischemic Stroke	8612 (3.30)	1095 (5.36)	774 (6.30)
Atrial Fibrillation (AF)	7187 (2.75)	931 (4.56)	778 (6.33)
Heart Failure (HF)	9107 (3.49)	1548 (7.58)	1157 (9.42)
Intracranial haemorrhage	3161 (1.19)	285 (1.40)	254 (2.07)

Ischemic Heart Disease (IHD)	20059 (7.68)	3474 (17.0)	1528 (12.4)
Osteoporosis	124 (0.047)	17 (0.083)	12 (0.098)
Hypertension	60321 (23.1)	7564 (37.0)	4472 (36.4)
Chronic Obstructive Pulmonary Disease	770 (0.295)	80 (0.392)	85 (0.692)

Table 27. Baseline characteristics for continuous variables in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death

Characteristics	Mean ± Standard Deviation		
	Total Cohort (n=261308)	Acute Myocardial Infarction (n=20419)	Sudden Cardiac Death (n=12282)
Age	66.0 ± 11.8	71.6 ± 10.7	72.9 ± 10.6
Follow-up Duration (days)	3552 ± 1201	2949 ± 1239	2008 ± 1143
Diabetes Duration (years)	4.77 ± 2.29	8.74 ± 4.12	9.95 ± 3.11
Liver Function Test			
Alkaline Phosphatase (U/L)	79.8 ± 37.4	81.3 ± 33.7	86.3 ± 51.5
Alanine Aminotransferase (U/L)	25.8 ± 24.0	22.6 ± 19.8	22.6 ± 19.3
Total Protein (g/L)	74.3 ± 6.99	73.9 ± 7.24	73.1 ± 7.46
Albumin (g/L)	38.7 ± 5.39	38.0 ± 5.33	37.0 ± 5.61
Complete Blood Count			
Lymphocyte Count ($x10^9$/L)	1.88 ± 1.05	1.85 ± 0.78	1.77 ± 1.58
Neutrophil Count ($x10^9$/L)	5.33 ± 2.68	5.62 ± 2.76	5.70 ± 2.86
Haemoglobin Count ($x10^9$/L)	12.8 ± 4.29	12.4 ± 1.87	12.2 ± 1.94
Lipid Profile			
High Density Lipoprotein Cholesterol (HDL-C) (mmol/L)	1.20 ± 0.34	1.15 ± 0.33	1.17 ± 0.36
Low Density Lipoprotein Cholesterol (LDL-C) (mmol/L)	2.92 ± 0.88	2.93 ± 0.93	2.88 ± 0.93

Total Cholesterol (mmol/L)	4.84 ± 1.03	4.84 ± 1.10	4.73 ± 1.08
Triglyceride (mmol/L)	1.72 ± 1.36	1.83 ± 1.52	1.72 ± 1.38
Renal Function Test			
Creatinine (umol/L)	103 ± 92	128 ± 125	139 ± 152
Potassium (mmol/L)	4.22 ± 0.48	4.27 ± 0.51	4.24 ± 0.52
Sodium (mmol/L)	139 ± 3	139 ± 3	139 ± 3.54
Urea (mmol/L)	6.85 ± 4.04	8.24 ± 5.01	8.52 ± 5.61
Glycaemic Control			
Fasting Blood Glucose (mmol/L)	7.75 ± 2.60	8.21 ± 2.00	8.12 ± 2.08
HbA1c (%)	7.44 ± 1.45	7.88 ± 1.25	7.83 ± 1.31

6.3.2. Acute myocardial infarction prediction

A total of 20419 patients suffered from AMI (annualized rate: 7.37%/year) with an annual frequency of 0.536 ± 8.74 episodes. The significant univariate predictors are summarized in **Table 21**. The following parameters were identified as significant predictors on multivariate regression (n=34015; **Table 28**): 1) age (HR= 1.02, 95% CI= [1.02, 1.03], p < 0.0001) and male sex (HR= 1.07, 95% CI= [1.01, 1.14], p= 0.023); 2) baseline anaemia (HR= 1.18, 95% CI= [1.10, 1.27], p < 0.0001); 3) serum creatinine (HR= 1.00, 95% CI= [1.00, 1.00], p < 0.0001); 4) serum HDL-C (HR= 0.802, 95% CI= [0.732, 0.878], p < 0.0001) and triglyceride (HR= 1.04, 95% CI= [1.03, 1.05], p < 0.0001); 5) comorbidities: ophthalmological diabetic complication (HR= 1.35, 95% CI= [1.22, 1.51], p < 0.0001), PVD (HR= 1.53, 95% CI= [1.18, 1.97], p= 0.001), IHD (HR= 1.59, 95% CI= [1.48, 1.71], p < 0.0001), hypertension (HR= 1.16, 95% CI= [1.09, 1.24], p < 0.001); 6) mean HbA1c (HR= 1.16, 95% CI= [1.12, 1.19], p < 0.0001).

Table 28. Univariate predictors for acute myocardial infarction in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death

	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.04	[1.04, 1.05]	< 0.0001
Male	1.10	[1.07, 1.14]	< 0.0001
Diabetes Duration	1.12	[1.11, 1.12]	< 0.0001
Mean Fasting Blood Glucose	1.12	[1.10, 1.13]	< 0.0001
Mean HbA1c	1.17	[1.15, 1.18]	< 0.0001
Baseline Anaemia	2.24	[2.15, 2.33]	< 0.0001
Liver Function Test			
Total Protein	0.991	[0.987, 0.994]	< 0.0001
Albumin	0.976	[0.972, 0.980]	< 0.0001
Renal Function Test			
Creatinine	1.00	[1.00, 1.00]	< 0.0001
Lipid Profile			
High Density Lipoprotein Cholesterol (HDL-C) (mmol/L)	0.587	[0.555, 0.620]	< 0.0001
Low Density Lipoprotein Cholesterol (LDL-C) (mmol/L)	1.02	[0.992, 1.04]	0.185
Total Cholesterol (mmol/L)	0.999	[0.983, 1.02]	0.858
Triglyceride (mmol/L)	1.04	[1.03, 1.04]	< 0.0001
Comorbidity			
Renal Diabetic Complication	2.56	[2.36, 2.79]	< 0.0001
Ophthalmological Diabetic Complication	2.69	[2.48, 2.91]	< 0.0001
Neurological Diabetic Complication	2.46	[2.14, 2.84]	< 0.0001
Peripheral Vascular Disease	3.83	[3.06, 4.78]	< 0.0001
Ischemic Stroke	1.72	[1.61, 1.82]	< 0.0001

Atrial Fibrillation	1.75	[1.64, 1.86]	< 0.0001
Heart Failure	2.40	[2.28, 2.53]	< 0.0001
Ischemic Heart Disease	2.63	[2.53, 2.72]	< 0.0001
Osteoporosis	2.03	[1.98, 2.09]	0.016
Hypertension	1.67	[1.62, 1.72]	< 0.0001
Chronic Obstructive Pulmonary Disease	1.36	[1.09, 1.69]	0.007

Both HDL-C and mean HbA1c showed linear relationships with AMI risk (**Figures 13 and 14**). After identifying the multivariate predictors (**Table 29**) and adjusting for the multivariate HR of the included parameters (**Table 30**), a score-based system was developed to predict AMI (**Table 31**). On ROC analysis, the AMI score had an AUC of 0.666 (95% CI= [0.662, 0.669]; **Figure 15**).

Figure 14. The association between mean HbA1c and acute myocardial infarction

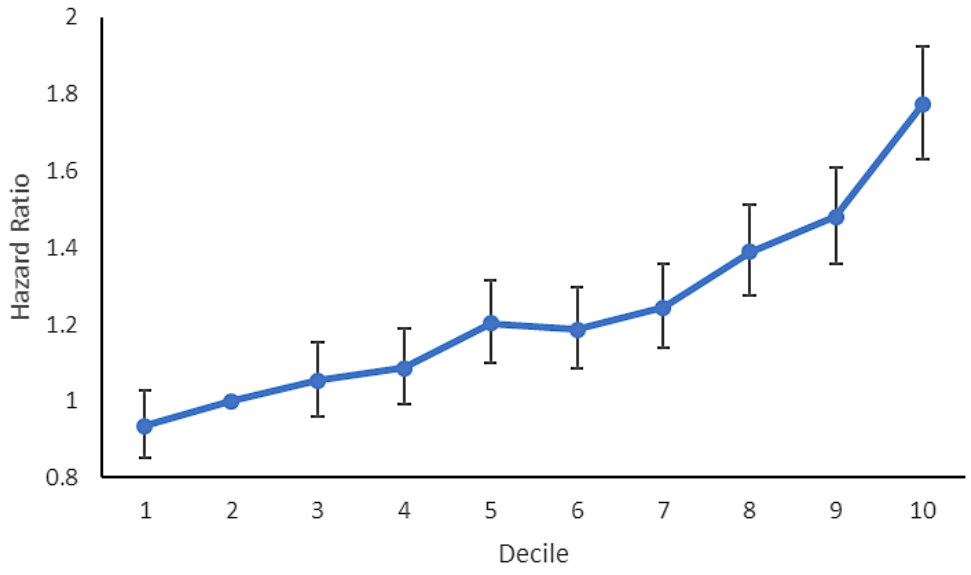


Figure 15. The association between mean high-density lipoprotein cholesterol and acute myocardial infarction

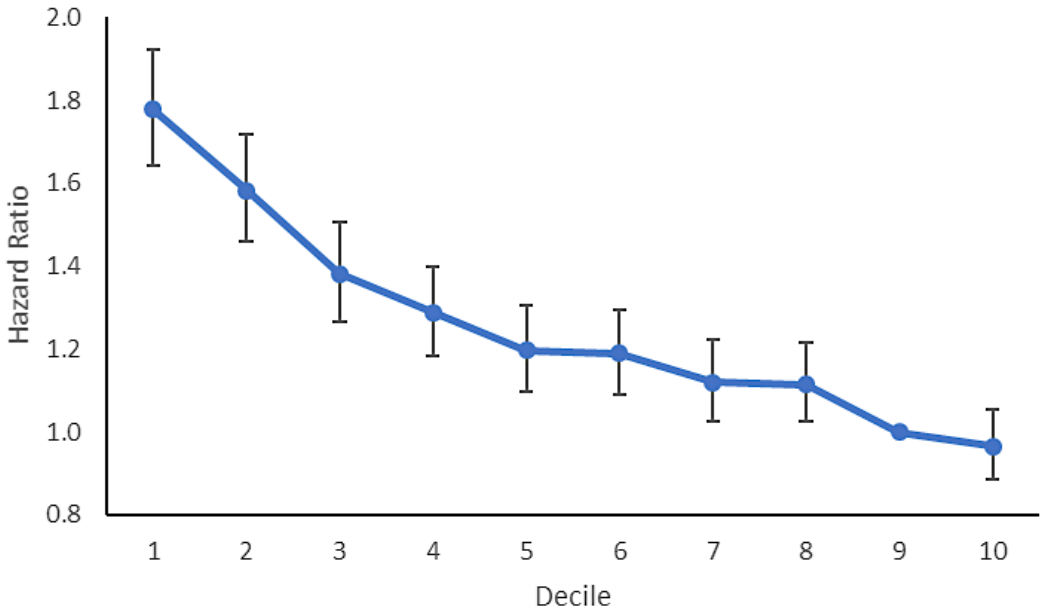


Table 29. Multivariate predictors for acute myocardial infarction in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death

	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.02	[1.02, 1.03]	< 0.0001
Male	1.07	[1.01, 1.14]	0.023
Mean Fasting Blood Glucose	0.994	[0.976, 1.01]	0.527
Mean HbA1c	1.16	[1.12, 1.19]	< 0.0001
Baseline Anaemia	1.18	[1.10, 1.27]	< 0.0001
Liver Function Test			
Total Protein	1.00	[0.996, 1.01]	0.651
Albumin	0.998	[0.991, 1.00]	0.523
Renal Function Test			
Creatinine	1.00	[1.00, 1.00]	< 0.0001
Lipid Profile			

High Density Lipoprotein Cholesterol (HDL-C)	0.802	[0.732, 0.878]	< 0.0001
Triglyceride	1.04	[1.03, 1.05]	< 0.0001
Comorbidity			
Renal Diabetic Complication	0.967	[0.865, 1.08]	0.561
Neurological Diabetic Complication	0.874	[0.732, 1.04]	0.132
Ophthalmological Diabetic Complication	1.35	[1.22, 1.51]	< 0.0001
Peripheral Vascular Disease	1.53	[1.18, 1.97]	0.001
Ischemic Stroke	0.991	[0.881, 1.11]	0.883
Atrial Fibrillation	0.962	[0.854, 1.08]	0.518
Heart Failure	1.01	[0.918, 1.11]	0.862
Ischemic Heart Disease	1.59	[1.48, 1.71]	< 0.0001
Osteoporosis	0.981	[0.542, 1.78]	0.950
Hypertension	1.16	[1.09, 1.24]	< 0.0001
Chronic Obstructive Pulmonary Disease	0.848	[0.539, 1.33]	0.473

Table 30. Multivariate hazard ratios of acute myocardial infarction predictive score parameters

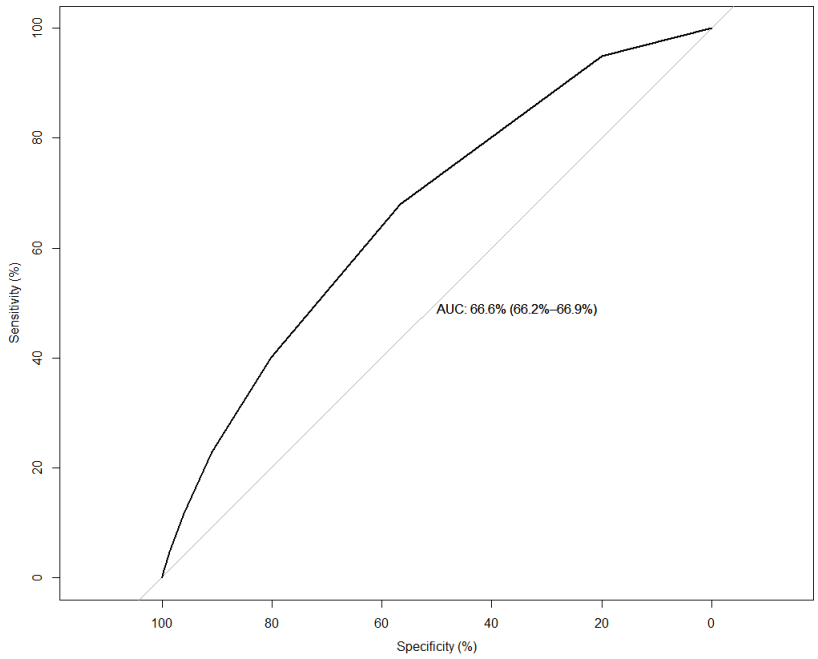
Citeria	Hazard Ratio
Age	1.04
Male	1.10
Baseline Anaemia	1.16
Creatinine	1.00
Mean HbA1c	1.19
High Density Lipoprotein Cholesterol (HDL-C)	0.776
Triglyceride	1.05
Ophthalmological Diabetic Complication	1.35
Peripheral Vascular Disease	1.38

Ischemic Heart Disease	1.64
Hypertension	1.24

Table 31. Acute myocardial infarction prediction score

Criteria	Cut-off	Score
Age (years)	>70	1
Sex	Male	1
Baseline Anaemia	Present	1
Creatinine (mmol/L)	>64.00	1
High Density Lipoprotein Cholesterol (mmol/L)	<1.07	1
Mean HbA1c (%)	> 8.51	1
Triglyceride (mmol/L)	>1.44	1
Ophthalmological Diabetic Complication	Present	1
Peripheral Vascular Disease	Present	1
Ischemic Heart Disease	Present	2
Hypertension	Present	1

Figure 16. The receiver operator characteristic curve for the acute myocardial infarction predictive score



6.3.3. Sudden cardiac death prediction

For risk stratification of SCD, 0.822% (n=2 149) patients were excluded because of AMI occurring before the SCD episode, or the SCD was associated with AMI. For this excluded subset of patients, only triglyceride levels were predictive of SCD. For the remainder of the cohort, SCD occurred in 12282 patients (annualized rate: 4.40%/year) at an annual frequency of 0.169 ± 0.569 episodes. Findings under univariate Cox regression are summarized in **Table 32**.

Table 32. Univariate predictors for sudden cardiac death in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death

Predictors	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.06	[1.05, 1.06]	< 0.0001
Male	1.23	[1.18, 1.27]	< 0.0001

Diabetes Duration	1.21	[1.21, 1.22]	< 0.0001
Mean Fasting Blood Glucose	1.03	[1.01, 1.04]	< 0.001
Mean HbA1c	1.13	[1.11, 1.15]	< 0.0001
Baseline Anaemia	2.74	[2.61, 2.87]	< 0.0001
Liver Function Test			
Total Protein	0.975	[0.971, 0.979]	< 0.0001
Albumin	0.946	[0.941, 0.951]	< 0.0001
Lipid Profile			
High Density Lipoprotein Cholesterol	0.818	[0.763, 0.876]	< 0.0001
Low Density Lipoprotein Cholesterol	0.977	[0.947, 1.01]	0.162
Total Cholesterol	0.896	[0.877, 0.915]	< 0.0001
Triglyceride	1.00	[0.987, 1.02]	0.802
Renal Function Test			
Creatinine	1.00	[1.00, 1.00]	< 0.0001
Comorbidity			
Renal Diabetic Complication	2.94	[2.65, 3.25]	< 0.0001
Ophthalmological Diabetic Complication	2.69	[2.43, 2.98]	< 0.0001
Neurological Diabetic Complication	2.50	[2.08, 3.00]	< 0.0001
Peripheral Vascular Disease	2.55	[1.81, 3.59]	< 0.0001
Ischemic Stroke	2.03	[1.89, 2.19]	< 0.0001
Atrial Fibrillation	2.50	[2.32, 2.69]	< 0.0001
Heart Failure	3.07	[2.89, 3.26]	< 0.0001
Ischemic Heart Disease	1.76	[1.67, 1.86]	< 0.0001
Osteoporosis	2.13	[1.21, 3.75]	0.009
Hypertension	1.96	[1.89, 2.03]	< 0.0001
Chronic Obstructive Pulmonary Disease	2.46	[1.99, 3.04]	< 0.0001

Multivariate Cox regression (n=33423) then identified following significant predictors, which were incorporated into the predictive score (**Table 33**): 1) age (HR= 1.03, 95% CI= [1.02, 1.03], p < 0.0001) and male sex (HR= 1.34, 95% CI= [1.23, 1.45], p < 0.0001); 2) baseline anaemia (HR= 1.41, 95% CI= [1.29, 1.54], p < 0.0001); 3) serum albumin (95% CI= 0.973, 95% CI= [0.964, 0.981], p < 0.0001); 4) serum total cholesterol (HR= 1.04, 95% CI= [1.00, 1.08], p= 0.033); 5) serum creatinine (HR= 1.00, 95% CI= [1.00, 1.00], p < 0.0001); 5) comorbidities: ophthalmological diabetic complication (HR= 1.23, 95% CI=[1.07, 1.41], p= 0.004), AF (HR=1.31, 95% CI= [1.14, 1.50], p < 0.0001) and HF (HR= 1.19, 95% CI= [1.06, 1.33], p=0.003); 6) mean HbA1c (HR= 1.11, 95% CI= [1.07, 1.15], p < 0.001).

Table 33. Multivariate predictors for sudden cardiac death in predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death

Parameter	Hazard Ratio	95% Confidence Interval	P-Value
Age	1.03	[1.02, 1.03]	< 0.0001
Male	1.34	[1.23, 1.45]	< 0.0001
Mean Fasting Blood Glucose	0.995	[0.972, 1.02]	0.684
Mean HbA1c	1.11	[1.07, 1.15]	< 0.0001
Baseline Anaemia	1.41	[1.29, 1.54]	< 0.0001
Liver Function Test			
Total Protein	1.00	[0.994, 1.01]	0.878
Albumin	0.973	[0.964, 0.981]	< 0.0001
Lipid Profile			
High Density Lipoprotein Cholesterol	0.905	[0.808, 1.01]	0.082
Total Cholesterol	1.04	[1.00, 1.08]	0.033
Renal Function Test			
Creatinine	1.00	[1.00, 1.00]	< 0.0001

Comorbidity			
Renal Diabetic Complication	1.02	[0.890, 1.17]	0.756
Ophthalmological Diabetic Complication	1.23	[1.07, 1.41]	0.004
Neurological Diabetic Complication	0.862	[0.687, 1.08]	0.195
Peripheral Vascular Disease	0.874	[0.586, 1.30]	0.510
Ischemic Stroke	1.13	[0.981, 1.29]	0.092
Atrial Fibrillation	1.31	[1.14, 1.50]	< 0.0001
Heart Failure	1.19	[1.06, 1.33]	0.003
Ischemic Heart Disease	1.00	[0.909, 1.11]	0.956
Osteoporosis	1.45	[0.755, 2.80]	0.263
Hypertension	1.06	[0.980, 1.16]	0.138
Chronic Obstructive Pulmonary Disease	1.23	[0.791, 1.91]	0.359

Both mean HbA1c and total cholesterol demonstrated a J-shaped relationship with non-AMI related SCD (**Figures 16 and 17**). Therefore, the cut-offs for mean HbA1c and total cholesterol were adjusted accordingly. The multivariate HR that the marks assigned in the score are shown in **Table 34**. None of the variables had HRs beyond the ranges of 0.67-1.5. Details of the scoring system are shown in **Table 35**, with ROC analysis showing an AUC of 0.677 (95% CI= [0.673, 0.682]) (**Figure 18**).

Figure 17. The association between mean HbA1c and acute myocardial infarction

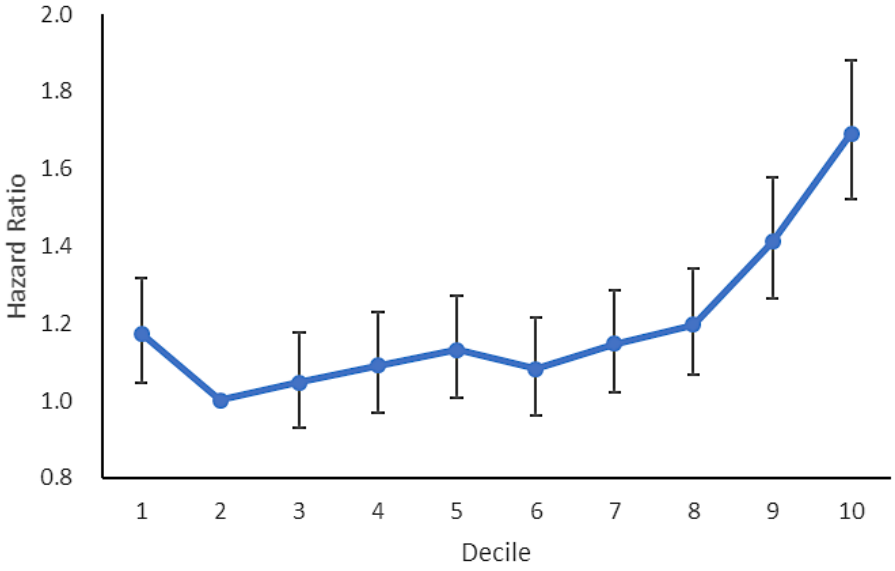


Figure 18. The association between mean total cholesterol and acute myocardial infarction

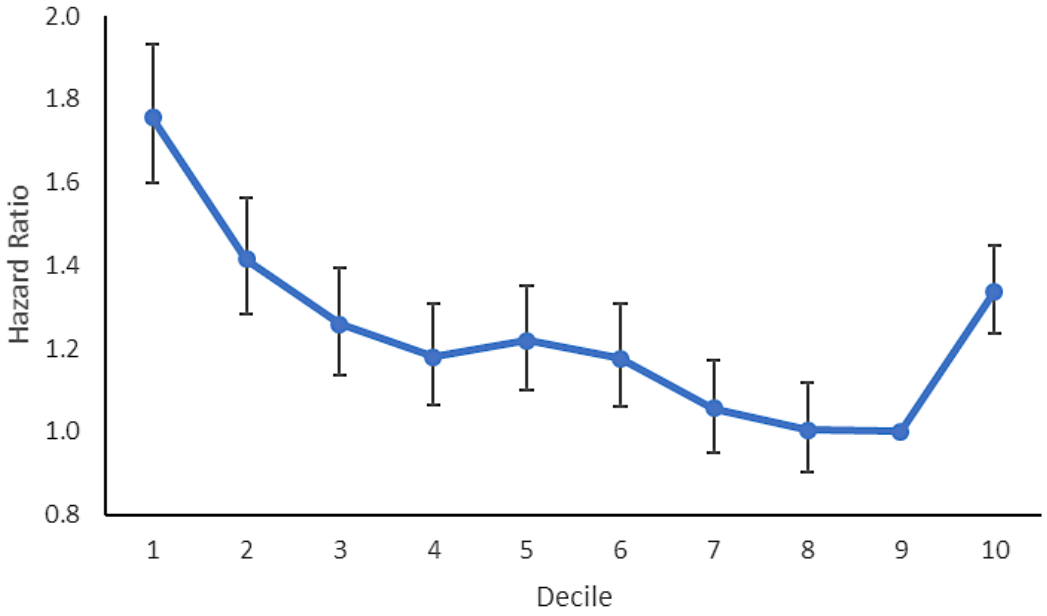


Figure 17 shows mean total cholesterol has a U-shaped relationship with the occurrence of acute myocardial infarction

Table 34. Multivariate hazard ratios of sudden cardiac death predictive score parameters

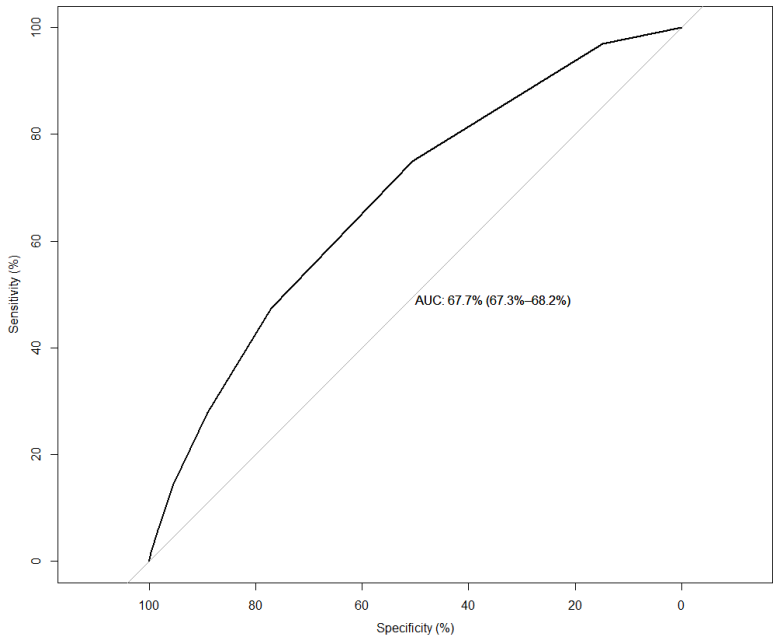
Criteria	Hazard Ratio
Male	1.30
Age	1.03

Baseline Anaemia	1.31
Albumin	0.972
Creatinine	1.00
Total Cholesterol	1.03
Mean HbA1c	1.10
Ophthalmological Diabetic Complication	1.24
Atrial Fibrillation	1.30
Heart Failure	1.23

Table 35. Sudden cardiac death predictive score

Criteria	Cut-off	Score
Age	>67	1
Sex	Male	1
Baseline Anaemia	Present	1
Total Cholesterol (mmol/L)	<5.00 or >6.11	1
Creatinine (mmol/L)	>93.6	1
Mean HbA1c (%)	<6.33 or > 7.79	1
Ophthalmological Diabetic Complication	Present	1
Atrial Fibrillation	Present	1
Heart Failure	Present	1

Figure 19. The receiver operator characteristic curve for sudden cardiac death predictive score



6.3.4. Machine learning survival analysis

A conditional inference survival forest (CISF) model was developed to predict AMI and SCD based on the baseline clinical variables. The optimal tree number of CISF model to predict AMI was set as 700 to predict AMI, while the number was set as 600 to predict SCD, based on the five-fold cross validation parameter selection results as shown in **Figures 19 and 20**.

Figure 20. Optimal tree number selection for conditional inference survival model (five-fold cross validation) to predict for acute myocardial infarction

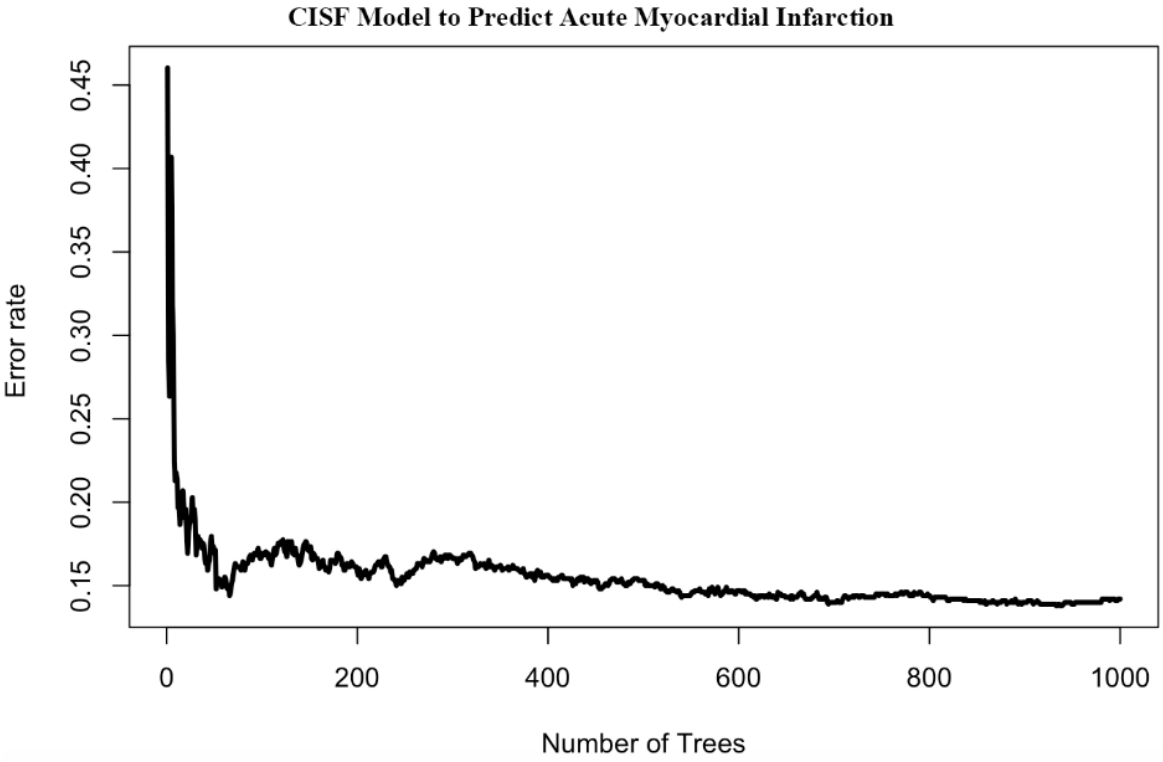
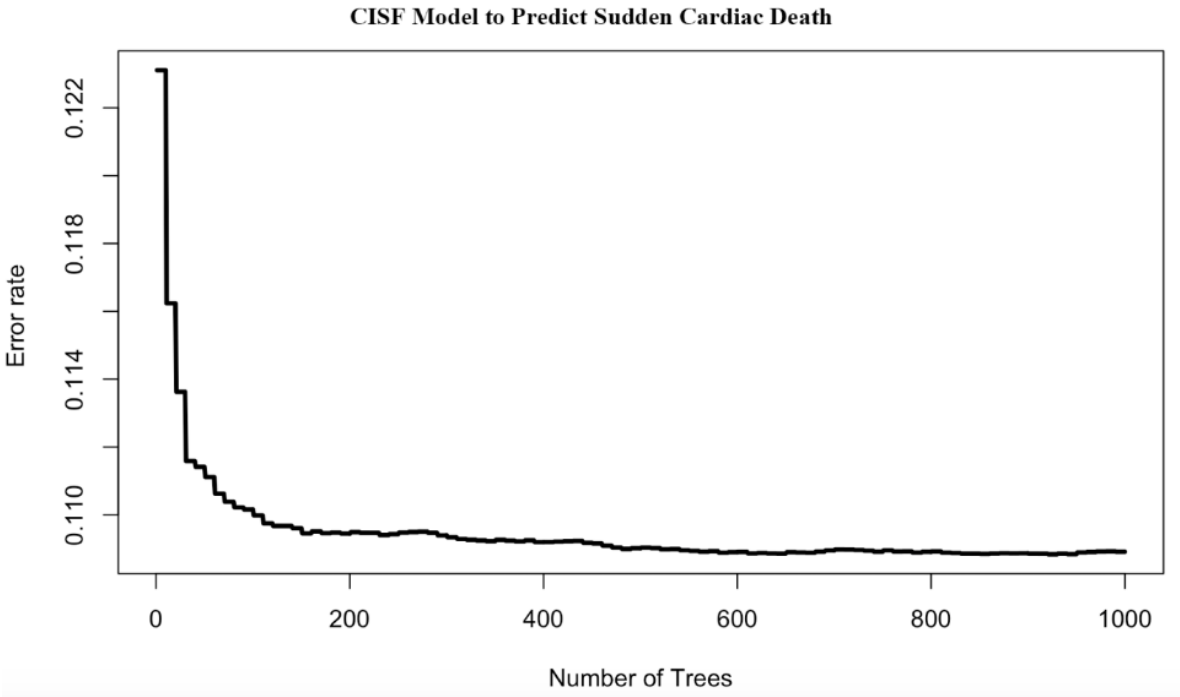


Figure 21. Optimal tree number selection for conditional inference survival forest model (five-fold cross validation) to predict for sudden cardiac death



Variable importance values and relative importance values of variables to predict AMI and non-AMI-related SCD are presented in **Table 36**. Creatinine and age were ranked as the most important predictors of AMI, followed by baseline anaemia, mean HbA1c, triglyceride, male sex, hypertension, and IHD (**Figure 21**). For non-AMI-related SCD, age and creatinine were the most important predictors, followed by baseline anaemia, mean HbA1c, HF, male sex, total cholesterol, AF, and ophthalmological diabetic complication (**Figure 22)**. The importance values of the different risk variables can be easily applied to construct predictive frailty scores of AMI and non-AMI-related SCD for clinical practice use.

Table 36. Variable importance ranking generated by conditional inference survival forest model

Acute Myocardial Infarction			**Sudden Cardiac Death**		
Variable	**Importance**	**Relative importance**	**Variable**	**Importance**	**Relative importance**
Creatinine (mmol/L)	0.1061	1.0000	Age, years	0.0986	1.0000
Age, years	0.0906	0.8545	Creatinine (mmol/L)	0.0923	0.9361
Baseline Anaemia	0.0156	0.1469	Baseline Anaemia	0.015	0.1517
Mean HbA1c (%)	0.0108	0.102	Mean HbA1c (%)	0.0126	0.1274
Triglyceride (mmol/L)	0.003	0.0284	Heart Failure	0.0119	0.1208
Male sex	0.0028	0.0268	Male sex	0.0086	0.0871
Hypertension	0.002	0.0193	Total Cholesterol (mmol/L)	0.0039	0.04
Ischemic Heart Disease	0.0012	0.011	Atrial Fibrillation	0.0024	0.0245
High Density Lipoprotein Cholesterol (mmol/L)	0.0005	0.0045	Ophthal-mological Diabetic Complication	0.0003	0.0032
Peripheral Vascular Disease	0.0001	0.0011			
Ophthal-mological	0.0000	0.0004			

Diabetic Complication					

Figure 22. Optimal tree number iteration and variable importance ranking generated by conditional inference survival forest model to predict acute myocardial infarction

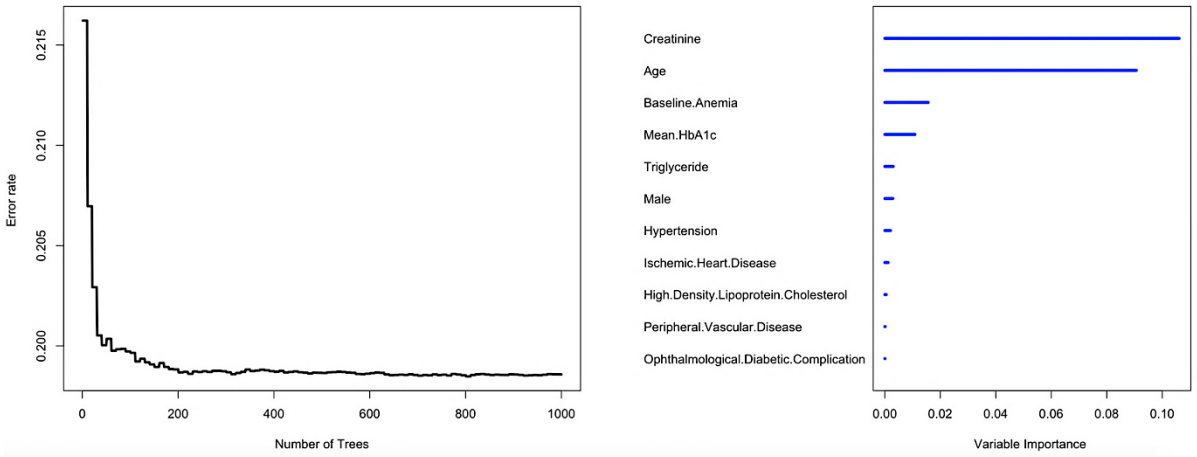
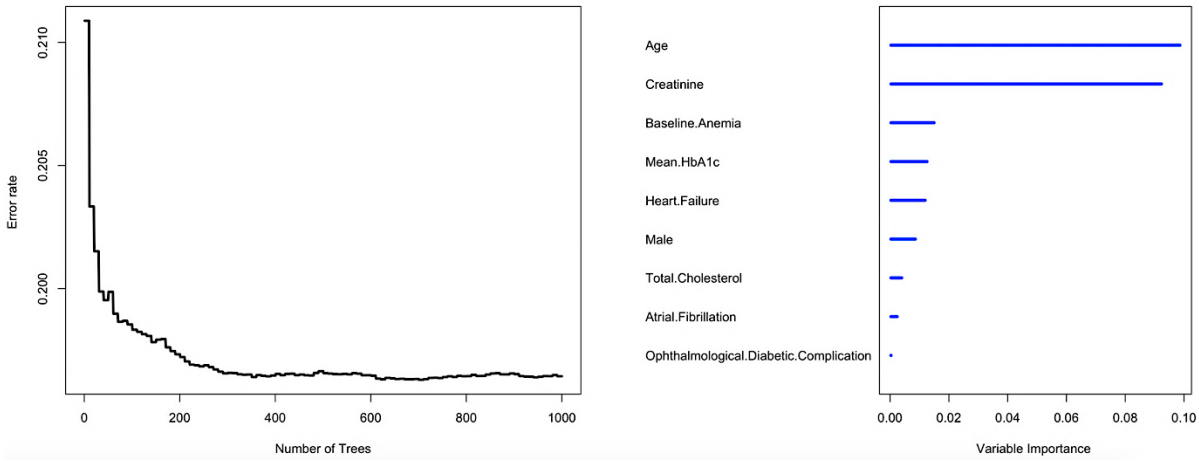


Figure 23. Optimal tree number iteration and variable importance ranking generated by conditional inference survival forest model to predict sudden cardiac death



The performance of the CISF model was compared with that of the RSF model and multivariate Cox for survival analysis (**Table 37**) using a five-fold cross validation approach. CISF model significantly improves the survival performance of AMI (precision: 0.91, recall: 0.89, AUC: 0.93, C-index: 0.91) and non-AMI related SCD (precision: 0.91, recall: 0.89, AUC: 0.89, C-index: 0.89) than RSF model and multivariate cox model.

Table 37. Comparisons between conditional inference survival forest (CISF), multivariate Cox and random survival forest (RSF) model (five-fold cross validation)

Outcome	Acute Myocardial Infarction				Sudden Cardiac Death			
Evaluators	Precision	Recall	AUC	C-Index	Precision	Recall	AUC	C-Index
CISF	0.9083	0.8851	0.9270	0.9029	0.9137	0.8900	0.8912	0.8918
RSF	0.8634	0.8606	0.8506	0.8290	0.8464	0.8406	0.8691	0.8536
Multivariate Cox	0.8197	0.7568	0.7255	0.7684	0.7918	0.8276	0.7412	0.8193

6.4. Discussion and limitations

6.4.1. Discussion

There are several major findings from the present study: 1) a combination of clinical and laboratory parameters can be used to predict AMI and SCD amongst patients with T2DM; 2) J/U-shaped relationships were not presented consistently across different cardiovascular adverse outcomes; 3) the J/ U-shaped relationships between mean HbA1c, HDL-C, and total cholesterol and adverse cardiovascular outcomes can be incorporated into scores for clinical risk stratification; 4) CISF model identified that albumin, age, creatinine, total protein, baseline anaemia, HF, and male gender are the most important predictors of both incident AMI and non-AMI related SCD, followed by hypertension, AF, HDL-C, mean FBG for AMI while mean FBG, hypertension, and mean HbA1c for non-AMI SCD;5) CISF significantly improves prediction performance of incident AMI and non-AMI SCD than RSF and multivariate Cox models.

Over recent years, there have been increasing reports on the J/U-shaped relationship between both glycaemic and cholesterol indices and diabetic adverse outcomes. However, these studies mostly focused on composite outcomes, such as all-cause mortality and major cardiovascular adverse events. (60, 181, 233, 234, 235) Currently, there is a lack of studies looking at the relationship between HbA1c and cholesterol indices with specific cardiovascular

adverse outcomes, such as AMI and SCD. In the present study, a linear relationship was observed between both mean HbA1c and HDL-C against AMI, whilst a J-shaped relationship was depicted for both mean HbA1c and total cholesterol against SCD. The incorporation of these biochemical variables into the risk scores yields comparable AUC to recent predictive models that involve machine learning techniques to account for latent interactions thus demonstrating the importance of involving biochemical indices in risk stratification. (236)

The difference that the mean HbA1c has against AMI and SCD can be explained by the different underlying pathogenic mechanisms. The linear relationship between mean HbA1c and AMI was supported by other studies with cohorts like the present study, comprised of younger patients with more diverse pre-existing macrovascular complications, which demonstrates the importance of personalised glycaemic control. (98, 237) Furthermore, coronary atherosclerosis is associated with insulin resistance, which also supports the linear relationship. (238, 239) In the DEVOTE trial, whilst hypoglycaemia increased the risk of cardiovascular diseases, the elevation in risk for non-fatal AMI and unstable angina was insignificant. These findings were consistent with the present study, where low mean HbA1c is associated with an increased risk for SCD but not AMI. (13)

On a separate note, the U-shaped relationship between mean HbA1c and SCD may be explained by the increased arrhythmic potential during both persistent hyperglycaemia and hypoglycaemia. Under chronic hyperglycaemia, persistently increased activation of calcium channels, and increased oxidative stress can induce arrhythmogenesis. (81, 82, 240, 241) By contrast, hypoglycaemia is a well-known trigger for ventricular tachyarrhythmia and is associated with delayed repolarization and altered repolarization gradients. (242, 243, 244) During prolonged hypoglycaemia, vagal reactivation occurs and the relative bradycardia increases the risk of atrial ectopy. (245) Severe hypoglycaemia was reported to increase the risk of arrhythmic death by 77% in the ORIGIN trial, which agrees with our findings. (246)

However, it should be noted that the J-shaped relationship is mainly attributed to the lowest decile of HbA1c, suggesting that the relationship may be disrupted by extreme cases of persistent hypoglycaemia. For patients with HbA1c values within the normal range, the relationship between HbA1c and SCD was linear.

The inverse relationship between HDL-C level and cardiovascular adverse outcomes is well established, and reinforced by recent findings of the inversed association between high lipoprotein function and atherosclerotic burden. (247) Recent studies exploring the relationship between cholesterol indices and cardiovascular events demonstrate that the J-shaped relationship is mainly present in LDL-C. (113, 235) The U-shaped relationship between HDL-C and all-cause mortality reported may be attributed to other causes of death, such as infection and external causes, and confounded by alcoholism which raises HDL. (248, 249, 250) These findings suggest that the J-shaped relationship between total cholesterol and SCD may be driven by LDL-C, given the observed linear association between HDL-C and AMI. Given that the J-shaped relationship between total cholesterol and SCD is mainly attributed to the highest decile for total cholesterol, there is also a possibility that the increase in SCD risk may only occur in outliers with extremely high total cholesterol. (251) The varied pathogenesis underlying different cardiovascular adverse outcomes suggests that cause-specific analysis of the relationship between both glycaemic and cholesterol, and cardiovascular mortality, should be performed.

The Cox proportional hazards model has been widely used for right-censored time-to-event data analysis since it is convenient for its flexibility and simplicity. However, their use is not appropriate when the proportional hazards assumption is violated. Extensions to the Cox proportional hazards model were developed but often remained dependent on restrictive functions (e.g., Heaviside functions) that are difficult to construct and implement. RSF models, as extensions of classification and regression trees and random forests, have been identified as

alternative survival data analysis methods when the proportional hazard assumption is violated (108). RSF-based models have been applied to enhance risk stratification in different clinical settings, including diabetes (153, 252, 253, 254, 255). However, the RSF model has been criticized for its bias due to favouring covariates with many split points (256). In our study, the CISF model was used for time-to-event survival data analysis in predicting AMI and non-AMI SCD (231, 232), which were shown to show superior predictive performance compared to RSF and multivariate Cox models.

6.4.2. Limitations

Several limitations should be noted for the present study. First of all, given its observational, data-based nature, it is susceptible to under-coding and coding error, with an inability to establish causal relationships. In addition, the large number of patients included in the analysis drove the high statistical significance but low HR in some predictive parameters. Thus, the findings of these parameters may be driven by the statistical power of the analysis and may have limited clinical significance. Furthermore, the duration of diabetes was not adjusted for, given the possible competing variable of time from baseline to outcome onset. This is also to avoid interference of inaccuracy in diabetic duration because of a lack of data beyond a decade before baseline. Additionally, the effect of medications was not accounted for due to the potential drug-drug interactions and effect on the laboratory markers, which would greatly complicate the analysis. Finally, data on other cardiovascular health predictors, such as smoking status, alcohol use, and family history of cardiac conditions, were unavailable due to limitations of our administrative database of not converting them into structured data for extraction.

6.5. Conclusion

A holistic combination of demographic, clinical, and laboratory indices can be used for the risk stratification of patients with T2DM against AMI and SCD. Cause-specific analysis should be applied to further examine the relationship between both mean HbA1c and lipid parameters against different cardiovascular adverse outcomes. The application of machine-learning techniques can improve the sensitivity and specificity of risk prediction by identifying the latent interactions between risk variables.

Chapter 7. Recapitulation and discussion

7.1. Background

MACE, including AMI, HF and ischaemic stroke, are major contributors to morbidity and mortality amongst patients with T2DM. Over the past decades, there has been a shift towards a more patient-centred, individualised approach in the long-term treatment of T2DM. Consequently, new, personalised disease-monitoring parameters have been explored over the past decade. Besides the absolute value of glycaemic and lipid levels, the temporal changes of glycaemic and lipids were recently demonstrated to predict the cardiovascular disease burden amongst patients with T2DM. The introduction of temporal variability in glycaemic and lipid control as a disease monitoring parameter in T2DM will lead to a paradigm shift in the management of T2DM. The treatment targets are no longer limited to the laboratory values at the moment of follow-up but extended to the rate and extent of changes in glycaemic and lipid control.

Over the last decade, there has been increasing evidence demonstrating the prognostic value of glycaemic and lipid variability in the MACE risk amongst patients with T2DM (77, 257). Although the exact mechanisms remain unclear, there are several hypothesised mechanisms underlying the pathogenesis of increased MACE risk under increased glycaemic and lipid variability. It should be noted that HbA1c variability is a better representation of long-term glycaemic variation since it is not affected by short-term glycaemic changes due to diet and medication changes given that haemoglobin has an average lifespan of 100 days. In addition, whilst different methods to assess the temporal variability have been proposed, including standard deviation, coefficient of variation, and scores based on percentage changes in the absolute value of the laboratory markers, no variability marker of absolute superiority has yet been identified (167).

Increased long-term glycaemic and lipid variability is hypothesised to lead to endothelial dysfunction via an increase in oxidative stress (198). Frequent episodes of hypoglycaemia can induce the release of reactive oxygen species, thereby leading to increased oxidative stress and creating a state of chronic inflammation (142, 258). Furthermore, fluctuations in glucose levels are associated with QTc prolongation and increased QTc dispersion, which elevates the risk of ventricular tachyarrhythmia and the resultant sudden cardiac death (205). Hypoglycaemia itself is arrhythmogenic and can reduce the myocardial tolerance to ischaemia and reperfusion injuries (80, 215). By contrast, chronic hyperglycaemia results in structural remodelling of the atria and ventricles, which is also arrhythmogenic and therefore increases the risk of MACE (81). Similarly, fluctuations in lipid levels also result in the release of atherogenic substances from unstable plaques, therefore, leading to an increase in oxidative stress (257).

Although it is well established that patients with poor glycaemic and lipid control have an increased risk for MACE, the demonstration of a high glycaemic variability as a risk factor helps to identify a different group of patients at an increased risk for cardiovascular events- those with marginal or adequate glycaemic control, but of high glycaemic/ lipid variability. The use of glycaemic/lipid variability over long term follow up alerts clinicians that fluctuating glycaemic may be a red flag for future cardiovascular events, where additional risk factor management would be needed to improve patient morbidity and mortality. Since HbA1c and lipid profiles are routinely measured upon follow up, a simple calculation, or automatically generated value of the respective SDs in patients with at least three measurements would be sufficient to alert clinicians for the increase in temporal variability. Future studies to identify cutoffs for the respective variability measures, such as SD or CV, will facilitate the clinical implementation of glycaemic/ lipid variability.

7.2. Glycaemic variability in the prediction of cardiovascular complications

Over the last decade, both long and short term glycaemic variability have gained academic interest in their potential use in cardiovascular risk prediction. (259) Short term glycaemic variability refers to glycaemic variation during the day and may be assessed through continuous glucose monitoring. SD, which is the rate of dispersion from the average glucose level, and CV, which is the SD of glucose divided by the average glucose level, are the two most common metrics used and recommended by the International Consensus on Use of Continuous Glucose Monitoring (260). Short term glycaemic variability reflects the presence and extent of change in glycaemic level, thus reflecting the stability of glycaemic control. Since sudden and large changes in glucose levels may precipitate adverse cardiovascular events such as arrhythmias and acute coronary syndrome, the use of short-term glycaemic variability may be predictive of major cardiovascular adverse events. Basic studies have shown that fluctuations in glycaemic levels increase the production of reactive oxygen species and lead to vascular damage. (261) These studies provide evidence that not only the average glycaemic control, but the consistency of optimal glucose control, are important in the control of cardiovascular risks. (262)

However, given the limited availability of continuous blood glucose monitoring, long term glycaemic variability is more commonly used in clinical practice and research. Long-term glycaemic variability refers to visit-to-visit fluctuations in glycaemic control, and may be reflected by the measuring of HbA1c and fasting/ postprandial blood glucose levels. Since HbA1c reflects blood glucose control over the past three months, it is a reliable indicator of the average glucose level. Clinical trials have consistently shown that increased long term glycaemic variability is associated with an increased risk of cardiovascular disease. A secondary analysis of the VADT shows that increased fasting glucose variability raises the risk of cardiovascular adverse events, particularly amongst those under intense glucose control.

(263) Interestingly, HbA1c variability was not found to be associated with MACE in this study. However, a post-hoc analysis of the ACCORD trial demonstrated that both HbA1c and fasting glucose variability were predictors for HF and cardiovascular disease, independent of changes in HbA1c, and variability in blood pressure, LDL-C, and hypoglycaemic events. (264) Although the underlying pathogenic mechanisms remain unclear, there is consistent evidence to support that increase in both long- and short-term glycaemic variability increases the risk of MACE.

The relationship between high HbA1c variability and MACE may extend to other diseases such as cancer (265). The pro-inflammatory state and increased oxidative stress environment produced by fluctuating levels of glycaemia are associated with increased risks of MACE, which may be partly mediated through endothelial dysfunction (133, 266). Furthermore, preclinical evidence shows that high glycaemic fluctuation also induces tissue oxidative stress, and can lead to increased apoptosis of cardiomyocytes (267).

The studies described in the present thesis thus provide further support that glycaemic variability may predict MACE risk amongst patients with T2DM, and it may be applied clinically in the form of clinical risk scores. Whilst existing studies have shown that glycaemic variability is associated with MACE in T2DM, the studies by our team have shown that high HbA1c variability is particularly associated with an increased risk of AF and AMI (268). Examining the association between HbA1c variability and specific MACEs provides further evidence that unstable glycaemic control contributes to an increased cardiovascular risk. Furthermore, the association between glycaemic variability and different MACE suggests that glycaemic variability may contribute to the occurrence of MACE through different mechanisms. For example, the occurrence of AF under high glycaemic fluctuations may be explained by the promotion of reactive oxygen species production, resulting in increased cardiac fibrosis and autonomic neuropathy (269). Catecholamine surge and sympathetic

activation under acute hypoglycaemic may also perpetuate AF (269). By contrast, the increased oxidative stress under a state of high glycaemic fluctuation is hypothesised to increase the risk of AMI by directly or indirectly causing endothelial dysfunction and the resulting accelerated atherosclerosis (270, 271).

In addition, temporal variability of glycaemic control is shown to be an independent predictor for MACE, regardless of the methodology of variability calculation or glycaemic measure examined. Therefore, variability measures that are easier to calculate and understand, such as standard deviation, may be a better option when it comes to applying glycaemic variability in the cardiovascular risk stratification models for patients with T2DM in the clinical setting.

Earlier studies on glycaemic variability in the risk of MACE amongst patients with T2DM studied Caucasian or Asian patients in general (96). The overwhelmingly Han Chinese demographic of the present registry provides strong evidence that glycaemic variability is a predictor of MACE amongst T2DM patients who are of Han Chinese descent. With the well-known ethnic or racial difference in T2DM incidence, showing that glycaemic variability predicts cardiovascular risk amongst Han Chinese patients with T2DM may provide specific targets to improve the disease prognosis of this specific patient population (272).

7.3. Lipid variability in the prediction of cardiovascular complication

Whilst T2DM patients were known to have an increased variability of plasma lipid level, it was only recently that its predictive values for cardiovascular risks were examined. c Prior studies have shown that high LDL-C, HDL-C and non-HDL-C variability were independent risk factors for cardiovascular disease in T2DM after adjusting for confounding variables (133, 273). In a study of over 125,000 patients with T2DM under primary care in Hong Kong, it was noted that lipid variability was a significant predictor for cardiovascular disease, with the

strongest predictors being LDL-C variability, particularly amongst the younger age group between 45-54 years old (274). Moreover, recent work has reported significant relationships between visit-to-visit cholesterol variability and long-term risks of new-onset HF and MACE also in patients under primary care in Hong Kong (275) as well as South Korea (131, 276).

Whilst the pathogenic mechanism remains unclear, it was hypothesised that increased LDL-C variability increases plaque instability due to the promotion of macrophage activation to form cholesterol core in atheroma, ultimately leading to plaque rupture (99). The atherogenic physiology under high lipid variability was further supported by the recent report of a positive association between the variability of LDL-C/ total cholesterol to HDL-C ratio and the percentage of coronary atherosclerotic plaque volume progression (124). However, clinical trials on the cardiovascular protective effects of interventions that target proteins related to lipid variability, such as cholesteryl ester transfer protein, remained negative (277). Therefore, further work is needed to elucidate the mechanisms underlying the increased cardiovascular risk amongst T2DM patients with high lipid variability.

By providing evidence that high lipid variability increases the risks for cardiovascular outcomes amongst T2DM patients, it uncovers the potential for mechanisms underlying lipid variability to be MACE-lowering therapeutic targets in T2DM. Besides reducing the absolute lipid level, interventions that stabilise lipid levels may be introduced to T2DM patients. A recent study reported single nucleotide polymorphisms (SNPs) associated with LDL-C and HDL-C variability. The SNPs associated with high LDL-C variability are related to apoprotein A5 (278). Apoprotein A5 increases lipoprotein lipase activity to facilitate the removal of lipoprotein from the circulation, hence its dysfunction results in lipid dysregulation with proatherogenic effects (279). The SNP associated with HDL-C variability is associated with *PXDNL*, a peroxidase homolog, which may promote atherosclerosis through increasing lipoprotein oxidation and impairing plasma lipid clearance (278). Interventions targeting

apoprotein A5, *PXDNL*, and other lipid variability-related proteins may have the potential to reduce cardiovascular risk among T2DM patients.

7.4. Novelty and implications in the use of machine learning in cardiovascular risk stratification

To address the lack of adjustments for confounders in the existing literature on the cardiovascular risk predictors amongst patients with T2DM, machine learning techniques were introduced to account for the interactions between glycaemic variability and other cardiovascular risk factors, therefore uncovering novel predictors and their complex inter-relationships. The present thesis provides the groundwork for machine learning-driven models of cardiovascular risk models in T2DM. By introducing the most relevant and influential predictors, the present thesis highlights the fundamental features to be included to ensure the accuracy of the cardiovascular risk stratification models. These features include biochemical parameters on glycaemic and lipid control, demographic factors such as age and sex, and different medical comorbidities. Models including genetic polymorphisms and lifestyle factors may be explored in the future. To maximize the cohort size for model generation, the Cox models were not validated against a training set. However, training sets were used in the validation of the machine learning models. The validity of the models can be examined by testing against other cohorts in e.g. United Kingdom and the United States. For example, the United Kingdom CALIBER dataset, with over 46 million patient, can be used for external validation (280).

Furthermore, the application of machine learning techniques improves the accuracy and precision of predictive models, ultimately tailoring the models to individual risk profiles. A recent study on more than 25,000 T2DM patients using insulin demonstrated that higher HbA1c and lipid variability increase the risk of peripheral vascular disease and mortality (281).

The application of regularised and weighted random survival forest models to account for interacting relationships improves the accuracy of prediction to a c-statistic of over 87%. Moreover, HbA1c and lipid variability were predictive of sudden cardiac death in patients with advanced stages of T2DM requiring insulin therapy (282). In an expanded cohort study of T2DM patients, a multivariable model incorporating indices of inflammation, HDL-C, total cholesterol, triglyceride, HbA1c and FBG, measures of variability of both HbA1c and FBG showed a c-statistic of 73%, which was improved to 86% and 87% using RSF and deep survival learning models, respectively (283). Machine learning algorithms have the potential to continuously learn and evolve with new data, therefore improving the risk models over time. With the ever-changing patient demographic, models that incorporate machine learning techniques can adapt to changes in the patient landscape, without the need for the generation of new models, therefore increasing the clinical applicability and sustainability of these models.

However, the application of machine-learning techniques also raises the difficulty for medical and non-medical professionals, without a data science background, to understand the predictive models, therefore may limit the applicability of these models. The present thesis attempted to address this issue by transforming the predictive models into weighted risk scores to improve the useability of the predictive models in a clinical setting. Thus, our models have now been piloted and tested in various clinics in Hong Kong, which have attracted positive feedback by not only patients but also clinicians. This is because of the availability of a simple, yet effective, individualised report, for each subject, who receive simple visualisation of their risks of different MACE events, accompanied by personalised recommendations and further support by the clinic nurse and physician.

In the future, the incorporation of different machine learning techniques can help to streamline the models and facilitate the communication of model findings. For example, the use of recursive feature elimination, where features were eliminated under repeated runs of the

importance-assigning algorithm can be applied to remove parameters that play an insignificant role in the risk prediction (284). Since the interactions between parameters have been accounted for in the importance-weighting of the features, it can ensure that no important interplaying feature has been eliminated. Additionally, visualisations, such as risk heatmaps, nomograms, and risk trajectories can help to communicate the calculated cardiovascular risk. The use of machine-learning driven visual aids can facilitate clinical discussion on treatment strategies and disease prognosis.

Whilst the present thesis illustrated the power of machine learning-incorporated models, data quality and standardisation remain significant obstacles to the integration of datasets from different centres and data sources for the generation of predictive models from diverse datasets. Currently, in countries such as the United Kingdom and the United States, there is a lack of a universal healthcare database. Patient information was kept in different electronic systems in different formats, sometimes even remaining in paper form. Besides the security and safety hazards, the absence of a universal system impairs the transparency and convenience of health information exchange (285). The accuracy and precision demonstrated in the present study show that the establishment of electronic healthcare databases with comprehensive, standardised data is not only beneficial for the record-keeping of patient information but is also a key to the successful application of machine learning techniques in models. Thus, this thesis showcases the power of leveraging the use of big data methods in analysing routinely collected health records, and the subsequent impact that can be made. Further implementation studies should be conducted in the future.

7.5. Limitations

Several limitations should be noted for the studies included in the present thesis. In terms of the dataset, the studies are developed from data from electronic health records, which

are susceptible to documentation errors and missing data. In addition, patients may be lost to follow-up, or have missed documentation of patient events due to the transfer of care to the private sector, where laboratory tests and patient records were not shared with the electronic healthcare database under the Hospital Authority. The electronic healthcare database was established in 1999, therefore comorbidities present before 1999 were not documented in the registry. Additionally, predictions in 10, 20, 30 years cannot be generated since the electronic healthcare database has only existed for less than 30 years. Furthermore, the inability of capture free text via CDARS can result in the underdiagnosis of conditions such as hypoglycaemia, therefore affecting the accuracy of prediction due to inherent variability in the dataset. Unfortunately, there is no access to free text. However, the quality of the dataset remains to be high in view of the highly complete dataset, with all laboratory results and drug prescriptions under the Hospital Authority integrated in CDARS. Clinical diagnoses are based on ICD codes entered by clinicians, and the database is linked to the death registry to ensure the completeness of the database. Furthermore, the registry mainly contains Han Chinese patients due to the demographic of the Hong Kong population, therefore limits the applicability of the findings to patients of other ethnicities.

In addition, there are limitations to the machine learning techniques applied. In general, machine learning models can be prone to overfitting, especially when there are many predictors in relativity to the sample size. Therefore, cross-validation and external validations with future studies are crucial to assess the generalisability of the model to new, unseen data (286). It is well established that machine learning models may be less interpretable than traditional statistical models. For example, in RSF models, the complex decision trees may make interpreting the specific contributions of individual variables challenging. In addition, data with unknown event times are censored for the processing of machine-learning models (286). The model may be sensitive to the extent and pattern of data censoring. However, patients with

unknown event times have been excluded in the present study, which limits the impact of the censoring mechanism on the studies in the present thesis.

In terms of limitations specific to survival forest models, they are unable to account for time-varying variables. Deep learning may be required for the analysis of time-varying variables (287). Moreover, it should be noted that the importance weighting assigned to different features in the machine learning model may not reflect the true clinical significance of the specific feature in the pathophysiological mechanism, hence should be interpreted with caution. The importance weighting of the feature may be influenced by the predictor strength, the correlation structure between predictors, and the censoring mechanism (288). A multidisciplinary approach with the collaboration of clinicians and data scientists can therefore ensure an appropriate interpretation and application of the machine learning models generated.

7.6. Future works

Future studies may be developed in several areas of the present topic. First of all, the present study focuses on the use of long-term glycaemic variability in the prediction of MACE given the limited availability of continuous blood glucose monitoring. The present thesis highlights the significant association between high glycaemic variability and increased cardiovascular risk, therefore promoting further research on the potential difference in effects between short- and long-term glycaemic variability. With the increasing application and improving accessibility of continuous blood glucose monitoring, studies on the involvement of short-term glycaemic variability in clinical models for the stratification of MACE risk amongst T2DM patients may be explored. Machine learning techniques may be applied to uncover the factors that affect one's short-term glycaemic variability, thus examining the possible difference in mechanisms underlying the effects on MACE between long- and short-term

glycaemic variability. Risk scores involving short term glycaemic variability may be generated, therefore improving the clinical applicability of the academic findings.

In addition, sudden episodes of clinically asymptomatic hypoglycaemia have been speculated to be one of the underlying mechanisms of increased MACE risk amongst patients with high glycaemic variability (209). The present thesis highlights the significant association between frequent hypoglycaemia and high glycaemic variability, thus providing further evidence for this hypothesis. Through the application of continuous blood glucose monitoring, episodes and extents of hypoglycaemia may be better captured, therefore able to evaluate the role of hypoglycaemia in the precipitation of MACE. With the combination of continuous ECG monitoring, the relationship between hypoglycaemia and arrhythmias, may be further evaluated and potentially establish a causative relationship. By marking patient events during the period of monitoring, the presence and types of symptoms present may be evaluated. Cause-specific analysis in future studies can also shed light on the potentially different mechanisms underlying high glycaemic/ lipid variability and different MACEs.

Furthermore, the present thesis has highlighted the association between glycaemic and lipid variability, hypoglycaemic frequency, and NLR, which provides a direction for future research in the pathophysiological mechanism underlying the increased cardiovascular risks in patients with high glycaemic/ lipid variability. The present thesis is the first to show an elevation in inflammatory markers under frequent hypoglycaemia in patients with high glycaemic/ lipid variability in a clinical setting, which will inspire future clinical studies to validate the present findings. External validations may be done through the examination of other inflammatory markers, the application of continuous blood glucose/ lipid monitoring, the exploration of short term glycaemic and lipid variability, the involvement of larger, more ethnically diverse cohorts, and more. The present findings also pave the way for basic studies to explore inflammatory pathways that may be activated under states of frequent glucose and

lipid level fluctuations. Substrates in the involved inflammatory pathways may be potential therapeutic targets to lower the cardiovascular risks of patients with T2DM. Other prognostic factors noted in the present study, sometimes ranking even more important than glycaemic and lipid variability, such as anaemia and age, may be explored in future studies on risk models.

In the future, deep learning and neural networks may be used in risk stratification models based on electronic healthcare records. The present thesis has set the grounds that machine learning is an effective method to improve the accuracy and precision of risk stratification models, in comparison to models generated through conventional methods, under the availability of high quality, standardized datasets. Other centres with electronic health databases available would be encouraged to explore machine learning techniques and collaborate with data scientists in future studies. It would also encourage a general shift to an electronic standardisation of healthcare data documentation for better consistency and accuracy.

Differing from machine learning, deep learning does not require human intervention in the correction of errors in its processing and analysis of data (289). Instead, neural networks were used to process training data for the development of algorithms. Through the use of deep learning, the generation and update of risk models can be automated, and time-varying variables may be included. For example, time-varying data from continuous blood/ ECG monitoring can be included in the risk models. A recent trial applying the use of deep learning in the optimisation of glycaemic control amongst T2DM patients has shown promising results (290). Deep learning has also been applied in the clustering of disease trajectories and prediction of glycaemic control in patients with T2DM (291, 292). Hence, the next step forward is to involve deep learning in the processing of electronic health records for the development of risk stratification models in T2DM patients.

7.7. Conclusion

To conclude, the present study demonstrated that glycaemic and lipid variability are both useful in the prediction of cardiovascular risk amongst patients with T2DM. The glycaemic and lipid variability are important for the evaluation of risks for specific MACE, such as SCD and AMI. The association between hypoglycaemic frequency, elevated inflammatory marker, glycaemic/ lipid variability, and MACE shed light on potential pathophysiological mechanisms involving inflammatory pathways.

In addition, machine learning techniques have been shown to improve clinical risk models, whilst maintaining the interpretability of the model. By accounting for complex interactions between variables, machine learning techniques can improve the accuracy and precision of the models. A multidisciplinary approach in the development of the risk models ensures the clinical applicability and interpretability of the risk stratification models.

In the future, preclinical studies furthering the potential inflammatory pathways triggered under hypoglycaemia may be performed. In addition, clinical studies with larger, more ethnically diverse cohorts and the involvement of continuous blood glucose and lipid monitoring can further elucidate the relationships between short-/long-term glycaemic and lipid variability and MACE in patients with T2DM. The application of machine and deep learning can be used in the prediction of specific MACE outcomes.

References

1. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9(th) edition. Diabetes Res Clin Pract. 2019;157:107843.
2. King P, Peacock I, Donnelly R. The UK prospective diabetes study (UKPDS): clinical and therapeutic implications for type 2 diabetes. Br J Clin Pharmacol. 1999;48(5):643-8.
3. Heller SR, Group AC. A summary of the ADVANCE Trial. Diabetes Care. 2009;32 Suppl 2:S357-61.
4. Currie CJ, Peters JR, Tynan A, Evans M, Heine RJ, Bracco OL, et al. Survival as a function of HbA(1c) in people with type 2 diabetes: a retrospective cohort study. Lancet. 2010;375(9713):481-9.
5. Huang ES, Davis AM. Glycemic Control in Older Adults With Diabetes Mellitus. JAMA. 2015;314(14):1509-10.
6. Anyanwagu U, Mamza J, Donnelly R, Idris I. Relationship between HbA1c and all-cause mortality in older patients with insulin-treated type 2 diabetes: results of a large UK Cohort Study. Age Ageing. 2019;48(2):235-40.
7. Arnold LW, Wang Z. The HbA1c and all-cause mortality relationship in patients with type 2 diabetes is J-shaped: a meta-analysis of observational studies. Rev Diabet Stud. 2014;11(2):138-52.
8. Gorst C, Kwok CS, Aslam S, Buchan I, Kontopantelis E, Myint PK, et al. Long-term Glycemic Variability and Risk of Adverse Outcomes: A Systematic Review and Meta-analysis. Diabetes Care. 2015;38(12):2354-69.
9. Forbes A, Murrells T, Mulnier H, Sinclair AJ. Mean HbA1c, HbA1c variability, and mortality in people with diabetes aged 70 years and older: a retrospective cohort study. The Lancet Diabetes & Endocrinology. 2018;6(6):476-86.
10. Sanchez J, Campuzano O, Iglesias A, Brugada R. Apunts Medicina de l'Esport. 2009;44(162):86-97.
11. Buse JB, Bigger JT, Byington RP, Cooper LS, Cushman WC, Friedewald WT, et al. Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: design and methods. Am J Cardiol. 2007;99(12a):21i-33i.
12. Wright RJ, Frier BM. Vascular disease and diabetes: is hypoglycaemia an aggravating factor? Diabetes Metab Res Rev. 2008;24(5):353-63.
13. Pieber TR, Marso SP, McGuire DK, Zinman B, Poulter NR, Emerson SS, et al. DEVOTE 3: temporal relationships between severe hypoglycaemia, cardiovascular outcomes and mortality. Diabetologia. 2018;61(1):58-65.
14. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, et al. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2022;45(11):2753-86.
15. Sfairopoulos D, Zhang N, Wang Y, Chen Z, Letsas KP, Tse G, et al. Association between sodium-glucose cotransporter-2 inhibitors and risk of sudden cardiac death or ventricular arrhythmias: a meta-analysis of randomized controlled trials. Europace. 2022;24(1):20-30.
16. Zhang N, Wang Y, Tse G, Korantzopoulos P, Letsas KP, Zhang Q, et al. Effect of sodium-glucose cotransporter-2 inhibitors on cardiac remodelling: a systematic review and meta-analysis. Eur J Prev Cardiol. 2022;28(17):1961-73.
17. McGuire DK, Shih WJ, Cosentino F, Charbonnel B, Cherney DZI, Dagogo-Jack S, et al. Association of SGLT2 Inhibitors With Cardiovascular and Kidney Outcomes in Patients With Type 2 Diabetes: A Meta-analysis. JAMA Cardiol. 2021;6(2):148-58.
18. Ni L, Yuan C, Chen G, Zhang C, Wu X. SGLT2i: beyond the glucose-lowering effect. Cardiovasc Diabetol. 2020;19(1):98.
19. Davies MJ, D'Alessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, et al. Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes

Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia. 2018;61(12):2461-98.
20. Kosiborod MN, Esterline R, Furtado RHM, Oscarsson J, Gasparyan SB, Koch GG, et al. Dapagliflozin in patients with cardiometabolic risk factors hospitalised with COVID-19 (DARE-19): a randomised, double-blind, placebo-controlled, phase 3 trial. Lancet Diabetes Endocrinol. 2021;9(9):586-94.
21. Chan JSK, Perone F, Bayatpoor Y, Tse G, Harky A. Emerging sodium-glucose cotransporter-2 inhibitor therapies for managing heart failure in patients with chronic kidney disease. Expert Opin Pharmacother. 2023:1-11.
22. Zhou J, Lee S, Leung KSK, Wai AKC, Liu T, Liu Y, et al. Incident heart failure and myocardial infarction in sodium-glucose cotransporter-2 vs. dipeptidyl peptidase-4 inhibitor users. ESC Heart Fail. 2022;9(2):1388-99.
23. Lee S, Zhou J, Leung KSK, Wai AKC, Jeevaratnam K, King E, et al. Comparison of Sodium-Glucose Cotransporter-2 Inhibitor and Dipeptidyl Peptidase-4 Inhibitor on the Risks of New-Onset Atrial Fibrillation, Stroke and Mortality in Diabetic Patients: A Propensity Score-Matched Study in Hong Kong. Cardiovasc Drugs Ther. 2022.
24. Mui JV, Li L, Chou OHI, Azfar N, Lee A, Hui J, et al. Comparing sodium-glucose cotransporter 2 inhibitors and dipeptidyl peptidase-4 inhibitors on new-onset depression: a propensity score-matched study in Hong Kong. Acta Diabetol. 2023.
25. Zhou J, Liu X, Chou OH, Li L, Lee S, Wong WT, et al. Lower risk of gout in sodium glucose cotransporter 2 (SGLT2) inhibitors versus dipeptidyl peptidase-4 (DPP4) inhibitors in type-2 diabetes. Rheumatology (Oxford). 2023;62(4):1501-10.
26. Tsampasian V, Elghazaly H, Chattopadhyay R, Ali O, Corballis N, Chousou PA, et al. Sodium glucose co-transporter 2 inhibitors in heart failure with preserved ejection fraction: a systematic review and meta-analysis. Eur J Prev Cardiol. 2022;29(6):e227-e9.
27. Tsampasian V, Baral R, Chattopadhyay R, Debski M, Joshi SS, Reinhold J, et al. The Role of SGLT2 Inhibitors in Heart Failure: A Systematic Review and Meta-Analysis. Cardiol Res Pract. 2021;2021:9927533.
28. Frias JP, Auerbach P, Bajaj HS, Fukushima Y, Lingvay I, Macura S, et al. Efficacy and safety of once-weekly semaglutide 2.0 mg versus 1.0 mg in patients with type 2 diabetes (SUSTAIN FORTE): a double-blind, randomised, phase 3B trial. Lancet Diabetes Endocrinol. 2021;9(9):563-74.
29. Frias JP, Bonora E, Nevarez Ruiz L, Li YG, Yu Z, Milicevic Z, et al. Efficacy and Safety of Dulaglutide 3.0 mg and 4.5 mg Versus Dulaglutide 1.5 mg in Metformin-Treated Patients With Type 2 Diabetes in a Randomized Controlled Trial (AWARD-11). Diabetes Care. 2021;44(3):765-73.
30. Buse JB, Wexler DJ, Tsapas A, Rossing P, Mingrone G, Mathieu C, et al. 2019 update to: Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetologia. 2020;63(2):221-8.
31. Group GSR, Nathan DM, Lachin JM, Balasubramanyam A, Burch HB, Buse JB, et al. Glycemia Reduction in Type 2 Diabetes - Glycemic Outcomes. N Engl J Med. 2022;387(12):1063-74.
32. Sattar N, McGuire DK, Pavo I, Weerakkody GJ, Nishiyama H, Wiese RJ, et al. Tirzepatide cardiovascular event risk assessment: a pre-specified meta-analysis. Nat Med. 2022;28(3):591-8.
33. Crisby M, Nordin-Fredriksson G, Shah PK, Yano J, Zhu J, Nilsson J. Pravastatin treatment increases collagen content and decreases lipid content, inflammation, metalloproteinases, and cell death in human carotid plaques: implications for plaque stabilization. Circulation. 2001;103(7):926-33.
34. Hattori K, Ozaki Y, Ismail TF, Okumura M, Naruse H, Kan S, et al. Impact of statin therapy on plaque characteristics as assessed by serial OCT, grayscale and integrated backscatter-IVUS. JACC Cardiovasc Imaging. 2012;5(2):169-77.

35. Lawler PR, Rosenson RS, Ko DT. Triglyceride reduction in secondary atherosclerotic cardiovascular disease prevention: core concepts in contemporary therapeutic targeting. Eur Heart J. 2020;41(15):1521-2.
36. Cholesterol Treatment Trialists C, Kearney PM, Blackwell L, Collins R, Keech A, Simes J, et al. Efficacy of cholesterol-lowering therapy in 18,686 people with diabetes in 14 randomised trials of statins: a meta-analysis. Lancet. 2008;371(9607):117-25.
37. Shepherd J, Barter P, Carmena R, Deedwania P, Fruchart JC, Haffner S, et al. Effect of lowering LDL cholesterol substantially below currently recommended levels in patients with coronary heart disease and diabetes: the Treating to New Targets (TNT) study. Diabetes Care. 2006;29(6):1220-6.
38. Sattar N, Preiss D, Murray HM, Welsh P, Buckley BM, de Craen AJ, et al. Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. Lancet. 2010;375(9716):735-42.
39. Cannon CP, Blazing MA, Giugliano RP, McCagg A, White JA, Theroux P, et al. Ezetimibe Added to Statin Therapy after Acute Coronary Syndromes. N Engl J Med. 2015;372(25):2387-97.
40. Stroes ESG, Bays HE, Banach M, Catapano AL, Duell PB, Laufs U, et al. Bempedoic acid lowers high-sensitivity C-reactive protein and low-density lipoprotein cholesterol: Analysis of pooled data from four phase 3 clinical trials. Atherosclerosis. 2023;373:1-9.
41. Nissen SE, Lincoff AM, Brennan D, Ray KK, Mason D, Kastelein JJP, et al. Bempedoic Acid and Cardiovascular Outcomes in Statin-Intolerant Patients. N Engl J Med. 2023;388(15):1353-64.
42. Ray KK, Wright RS, Kallend D, Koenig W, Leiter LA, Raal FJ, et al. Two Phase 3 Trials of Inclisiran in Patients with Elevated LDL Cholesterol. N Engl J Med. 2020;382(16):1507-19.
43. Low Wang CC, Hess CN, Hiatt WR, Goldfine AB. Clinical Update: Cardiovascular Disease in Diabetes Mellitus: Atherosclerotic Cardiovascular Disease and Heart Failure in Type 2 Diabetes Mellitus - Mechanisms, Management, and Clinical Considerations. Circulation. 2016;133(24):2459-502.
44. Sabatine MS, Giugliano RP, Wiviott SD, Raal FJ, Blom DJ, Robinson J, et al. Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. N Engl J Med. 2015;372(16):1500-9.
45. Seidah NG, Garcon D. Expanding Biology of PCSK9: Roles in Atherosclerosis and Beyond. Curr Atheroscler Rep. 2022;24(10):821-30.
46. Liu X, Suo R, Chan CZY, Liu T, Tse G, Li G. The immune functions of PCSK9: Local and systemic perspectives. J Cell Physiol. 2019;234(11):19180-8.
47. Laufs U, Parhofer KG, Ginsberg HN, Hegele RA. Clinical review on triglycerides. Eur Heart J. 2020;41(1):99-109c.
48. Das Pradhan A, Glynn RJ, Fruchart JC, MacFadyen JG, Zaharris ES, Everett BM, et al. Triglyceride Lowering with Pemafibrate to Reduce Cardiovascular Risk. N Engl J Med. 2022;387(21):1923-34.
49. Bhatt DL, Steg PG, Miller M, Brinton EA, Jacobson TA, Ketchum SB, et al. Cardiovascular Risk Reduction with Icosapent Ethyl for Hypertriglyceridemia. N Engl J Med. 2019;380(1):11-22.
50. Nicholls SJ, Lincoff AM, Garcia M, Bash D, Ballantyne CM, Barter PJ, et al. Effect of High-Dose Omega-3 Fatty Acids vs Corn Oil on Major Adverse Cardiovascular Events in Patients at High Cardiovascular Risk: The STRENGTH Randomized Clinical Trial. Jama. 2020;324(22):2268-80.
51. Zinman B, Marso SP, Poulter NR, Emerson SS, Pieber TR, Pratley RE, et al. Day-to-day fasting glycaemic variability in DEVOTE: associations with severe hypoglycaemia and cardiovascular outcomes (DEVOTE 2). Diabetologia. 2018;61(1):48-57.
52. Hirakawa Y, Arima H, Zoungas S, Ninomiya T, Cooper M, Hamet P, et al. Impact of visit-to-visit glycemic variability on the risks of macrovascular and microvascular events and all-cause mortality in type 2 diabetes: the ADVANCE trial. Diabetes Care. 2014;37(8):2359-65.
53. Zhou JJ, Schwenke DC, Bahn G, Reaven P, Investigators V. Glycemic Variation and Cardiovascular Risk in the Veterans Affairs Diabetes Trial. Diabetes Care. 2018;41(10):2187-94.

54. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? PLoS One. 2017;12(4):e0174944.
55. Lee S, Liu T, Zhou J, Zhang Q, Wong WT, Tse G. Predictions of diabetes complications and mortality using hba1c variability: a 10-year observational cohort study. Acta Diabetol. 2021;58(2):171-80.
56. Li S, Hou L, Zhu S, Yi Q, Liu W, Zhao Y, et al. Lipid Variability and Risk of Cardiovascular Diseases and All-Cause Mortality: A Systematic Review and Meta-Analysis of Cohort Studies. Nutrients. 2022;14(12).
57. Wan EYF, Yu EYT, Chin WY, Barrett JK, Mok AHY, Lau CST, et al. Greater variability in lipid measurements associated with cardiovascular disease and mortality: 10-year diabetes cohort study. Diabetes Obes Metab. 2020.
58. Group AS, Buse JB, Bigger JT, Byington RP, Cooper LS, Cushman WC, et al. Action to Control Cardiovascular Risk in Diabetes (ACCORD) trial: design and methods. Am J Cardiol. 2007;99(12A):21i-33i.
59. American Diabetes A. 12. Older Adults: Standards of Medical Care in Diabetes-2020. Diabetes Care. 2020;43(Suppl 1):S152-S62.
60. Forbes A, Murrells T, Mulnier H, Sinclair AJ. Mean HbA1c, HbA1c variability, and mortality in people with diabetes aged 70 years and older: a retrospective cohort study. Lancet Diabetes Endocrinol. 2018;6(6):476-86.
61. Li S, Nemeth I, Donnelly L, Hapca S, Zhou K, Pearson ER. Visit-to-Visit HbA1c Variability Is Associated With Cardiovascular Disease and Microvascular Complications in Patients With Newly Diagnosed Type 2 Diabetes. Diabetes Care. 2020;43(2):426-32.
62. Su JB, Zhao LH, Zhang XL, Cai HL, Huang HY, Xu F, et al. HbA1c variability and diabetic peripheral neuropathy in type 2 diabetic patients. Cardiovasc Diabetol. 2018;17(1):47.
63. Luk AO, Ma RC, Lau ES, Yang X, Lau WW, Yu LW, et al. Risk association of HbA1c variability with chronic kidney disease and cardiovascular disease in type 2 diabetes: prospective analysis of the Hong Kong Diabetes Registry. Diabetes Metab Res Rev. 2013;29(5):384-90.
64. Waden J, Forsblom C, Thorn LM, Gordin D, Saraheimo M, Groop PH, et al. A1C variability predicts incident cardiovascular events, microalbuminuria, and overt diabetic nephropathy in patients with type 1 diabetes. Diabetes. 2009;58(11):2649-55.
65. Kahal H, Halama A, Aburima A, Bhagwat AM, Butler AE, Grauman J, et al. Effect of induced hypoglycemia on inflammation and oxidative stress in type 2 diabetes and control subjects. Sci Rep. 2020;10(1):4750.
66. Wang J, Alexanian A, Ying R, Kizhakekuttu TJ, Dharmashankar K, Vasquez-Vivar J, et al. Acute exposure to low glucose rapidly induces endothelial dysfunction and mitochondrial oxidative stress: role for AMP kinase. Arterioscler Thromb Vasc Biol. 2012;32(3):712-20.
67. Li CK, Xu Z, Ho J, Lakhani I, Liu YZ, Bazoukis G, et al. Association of NPAC score with survival after acute myocardial infarction. Atherosclerosis. 2020;301:30-6.
68. Ju C, Lai RWC, Li KHC, Hung JKF, Lai JCL, Ho J, et al. Comparative cardiovascular risk in users versus non-users of xanthine oxidase inhibitors and febuxostat versus allopurinol users. Rheumatology (Oxford). 2019.
69. Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34). UK Prospective Diabetes Study (UKPDS) Group. Lancet. 1998;352(9131):854-65.
70. Andersson C, van Gaal L, Caterson ID, Weeke P, James WP, Coutinho W, et al. Relationship between HbA1c levels and risk of cardiovascular adverse outcomes and all-cause mortality in overweight and obese cardiovascular high-risk women and men with type 2 diabetes. Diabetologia. 2012;55(9):2348-55.
71. Action to Control Cardiovascular Risk in Diabetes Study G, Gerstein HC, Miller ME, Byington RP, Goff DC, Jr., Bigger JT, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med. 2008;358(24):2545-59.

72. Li W, Katzmarzyk PT, Horswell R, Wang Y, Johnson J, Hu G. HbA1c and all-cause mortality risk among patients with type 2 diabetes. Int J Cardiol. 2016;202:490-6.
73. Carson AP, Fox CS, McGuire DK, Levitan EB, Laclaustra M, Mann DM, et al. Low hemoglobin A1c and risk of all-cause mortality among US adults without diabetes. Circ Cardiovasc Qual Outcomes. 2010;3(6):661-7.
74. Crane PK, Walker R, Hubbard RA, Li G, Nathan DM, Zheng H, et al. Glucose levels and risk of dementia. N Engl J Med. 2013;369(6):540-8.
75. Rogers SC, Zhang X, Azhar G, Luo S, Wei JY. Exposure to high or low glucose levels accelerates the appearance of markers of endothelial cell senescence and induces dysregulation of nitric oxide synthase. J Gerontol A Biol Sci Med Sci. 2013;68(12):1469-81.
76. Zaslavsky O, Walker RL, Crane PK, Gray SL, Larson EB. Glucose Levels and Risk of Frailty. J Gerontol A Biol Sci Med Sci. 2016;71(9):1223-9.
77. Gu J, Pan JA, Fan YQ, Zhang HL, Zhang JF, Wang CQ. Prognostic impact of HbA1c variability on long-term outcomes in patients with heart failure and type 2 diabetes mellitus. Cardiovasc Diabetol. 2018;17(1):96.
78. Cheng D, Fei Y, Liu Y, Li J, Xue Q, Wang X, et al. HbA1C variability and the risk of renal status progression in Diabetes Mellitus: a meta-analysis. PLoS One. 2014;9(12):e115509.
79. Monnier L, Mas E, Ginet C, Michel F, Villon L, Cristol JP, et al. Activation of oxidative stress by acute glucose fluctuations compared with sustained chronic hyperglycemia in patients with type 2 diabetes. JAMA. 2006;295(14):1681-7.
80. Costantino S, Paneni F, Battista R, Castello L, Capretti G, Chiandotto S, et al. Impact of Glycemic Variability on Chromatin Remodeling, Oxidative Stress, and Endothelial Dysfunction in Patients With Type 2 Diabetes and With Target HbA1c Levels. Diabetes. 2017;66(9):2472-82.
81. Tse G, Lai ET, Tse V, Yeo JM. Molecular and Electrophysiological Mechanisms Underlying Cardiac Arrhythmogenesis in Diabetes Mellitus. J Diabetes Res. 2016;2016:2848759.
82. Tse G, Yan BP, Chan YW, Tian XY, Huang Y. Reactive Oxygen Species, Endoplasmic Reticulum Stress and Mitochondrial Dysfunction: The Link with Cardiac Arrhythmogenesis. Front Physiol. 2016;7:313.
83. Skriver MV, Sandbaek A, Kristensen JK, Stovring H. Relationship of HbA1c variability, absolute changes in HbA1c, and all-cause mortality in type 2 diabetes: a Danish population-based prospective observational study. BMJ Open Diabetes Res Care. 2015;3(1):e000060.
84. Fung ACH, Tse G, Cheng HL, Lau ESH, Luk A, Ozaki R, et al. Depressive Symptoms, Co-Morbidities, and Glycemic Control in Hong Kong Chinese Elderly Patients With Type 2 Diabetes Mellitus. Front Endocrinol (Lausanne). 2018;9:261.
85. Alwafi H, Wei L, Naser AY, Mongkhon P, Tse G, Man KKC, et al. Trends in oral anticoagulant prescribing in individuals with type 2 diabetes mellitus: a population-based study in the UK. BMJ Open. 2020;10(5):e034573.
86. Xiong Z, Liu T, Tse G, Gong M, Gladding PA, Smaill BH, et al. A Machine Learning Aided Systematic Review and Meta-Analysis of the Relative Risk of Atrial Fibrillation in Patients With Diabetes Mellitus. Front Physiol. 2018;9:835.
87. Bailey CJ, Aschner P, Del Prato S, LaSalle J, Ji L, Matthaei S, et al. Individualized glycaemic targets and pharmacotherapy in type 2 diabetes. Diab Vasc Dis Res. 2013;10(5):397-409.
88. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, et al. Management of hyperglycemia in type 2 diabetes: a patient-centered approach: position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care. 2012;35(6):1364-79.
89. Harris SB, Cheng AYY, Davies MJ, Gerstein HC, Green JB, Skolnik N. Person-Centered, Outcomes-Driven Treatment: A New Paradigm for Type 2 Diabetes in Primary Care. Arlington (VA)2020.

90. Rama Chandran S, R AV, Thomas A, Lim LL, Ratnasingam J, Tan A, et al. Role of Composite Glycemic Indices: A Comparison of the Comprehensive Glucose Pentagon Across Diabetes Types and HbA1c Levels. Diabetes Technol Ther. 2020;22(2):103-11.
91. Rama Chandran S, Tay WL, Lye WK, Lim LL, Ratnasingam J, Tan ATB, et al. Beyond HbA1c: Comparing Glycemic Variability and Glycemic Indices in Predicting Hypoglycemia in Type 1 and Type 2 Diabetes. Diabetes Technol Ther. 2018;20(5):353-62.
92. da Silva DE, Grande AJ, Roever L, Tse G, Liu T, Biondi-Zoccai G, et al. High-Intensity Interval Training in Patients with Type 2 Diabetes Mellitus: a Systematic Review. Curr Atheroscler Rep. 2019;21(2):8.
93. Lakhani I, Gong M, Wong WT, Bazoukis G, Lampropoulos K, Wong SH, et al. Fibroblast growth factor 21 in cardio-metabolic disorders: a systematic review and meta-analysis. Metabolism. 2018;83:11-7.
94. Qi W, Zhang N, Korantzopoulos P, Letsas KP, Cheng M, Di F, et al. Serum glycated hemoglobin level as a predictor of atrial fibrillation: A systematic review with meta-analysis and meta-regression. PLoS One. 2017;12(3):e0170955.
95. Ceriello A, De Cosmo S, Rossi MC, Lucisano G, Genovese S, Pontremoli R, et al. Variability in HbA1c, blood pressure, lipid parameters and serum uric acid, and risk of development of chronic kidney disease in type 2 diabetes. Diabetes Obes Metab. 2017;19(11):1570-8.
96. Echouffo-Tcheugui JB, Zhao S, Brock G, Matsouaka RA, Kline D, Joseph JJ. Visit-to-Visit Glycemic Variability and Risks of Cardiovascular Events and All-Cause Mortality: The ALLHAT Study. Diabetes Care. 2019;42(3):486-93.
97. Ghasemzadeh Z, Abdi H, Asgari S, Tohidi M, Khalili D, Valizadeh M, et al. Divergent pathway of lipid profile components for cardiovascular disease and mortality events: Results of over a decade follow-up among Iranian population. Nutr Metab (Lond). 2016;13:43.
98. Bardini G, Innocenti M, Rotella CM, Giannini S, Mannucci E. Variability of triglyceride levels and incidence of microalbuminuria in type 2 diabetes. J Clin Lipidol. 2016;10(1):109-15.
99. Bangalore S, Breazna A, DeMicco DA, Wun CC, Messerli FH, Committee TNTS, et al. Visit-to-visit low-density lipoprotein cholesterol variability and risk of cardiovascular outcomes: insights from the TNT trial. J Am Coll Cardiol. 2015;65(15):1539-48.
100. Bardini G, Dicembrini I, Rotella CM, Giannini S. Lipids seasonal variability in type 2 diabetes. Metabolism. 2012;61(12):1674-7.
101. Pineda A, Cubeddu LX. Statin rebound or withdrawal syndrome: does it exist? Curr Atheroscler Rep. 2011;13(1):23-30.
102. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. The annals of applied statistics. 2008;2(3):841-60.
103. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. Circ Cardiovasc Qual Outcomes. 2011;4(1):39-45.
104. Qi Y, Bar‐Joseph Z, Klein‐Seetharaman J. Evaluation of different biological data and computational classification methods for use in protein interaction prediction. Proteins: Structure, Function, and Bioinformatics. 2006;63(3):490-500.
105. Man KKC, Chan EW, Ip P, Coghill D, Simonoff E, Chan PKL, et al. Prenatal antidepressant use and risk of attention-deficit/hyperactivity disorder in offspring: population based cohort study. BMJ. 2017;357:j2350.
106. Law SWY, Lau WCY, Wong ICK, Lip GYH, Mok MT, Siu CW, et al. Sex-Based Differences in Outcomes of Oral Anticoagulation in Patients With Atrial Fibrillation. J Am Coll Cardiol. 2018;72(3):271-82.
107. Zhou J, Wang X, Lee S, Wu WKK, Cheung BMY, Zhang Q, et al. Proton pump inhibitor or famotidine use and severe COVID-19 disease: a propensity score-matched territory-wide study. Gut. 2020.

108. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat. 2008;2(3):841-60.
109. Breiman L. Random Forests. Machine Learning. 2001;45(1):5-32.
110. Ishwaran H. Variable importance in binary regression trees and forests. Electron J Statist. 2007;1:519-37.
111. Steele AJ, Denaxas SC, Shah AD, Hemingway H, Luscombe NM. Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease. PLOS ONE. 2018;13(8):e0202344.
112. Lee S, Zhou J, Leung KSK, Wu WKK, Wong WT, Liu T, et al. Development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong. BMJ Open Diabetes Res Care. 2021.
113. Liu L, Shen G, Huang JY, Yu YL, Chen CL, Huang YQ, et al. U-shaped association between low-density lipid cholesterol and diabetes mellitus in patients with hypertension. Lipids Health Dis. 2019;18(1):163.
114. Li X, Guan B, Wang Y, Tse G, Zou F, Khalid BW, et al. Association between high-density lipoprotein cholesterol and all-cause mortality in the general population of northern China. Sci Rep. 2019;9(1):14426.
115. Yi SW, Yi JJ, Ohrr H. Total cholesterol and all-cause mortality by sex and age: a prospective cohort study among 12.8 million adults. Sci Rep. 2019;9(1):1596.
116. Ahmadi SF, Streja E, Zahmatkesh G, Streja D, Kashyap M, Moradi H, et al. Reverse Epidemiology of Traditional Cardiovascular Risk Factors in the Geriatric Population. J Am Med Dir Assoc. 2015;16(11):933-9.
117. Lavie G, Hoshen M, Leibowitz M, Benis A, Akriv A, Balicer R, et al. Statin Therapy for Primary Prevention in the Elderly and Its Association with New-Onset Diabetes, Cardiovascular Events, and All-Cause Mortality. Am J Med. 2021;134(5):643-52.
118. Pilotto A, Panza F, Copetti M, Simonato M, Sancarlo D, Gallina P, et al. Statin Treatment and Mortality in Community-Dwelling Frail Older Patients with Diabetes Mellitus: A Retrospective Observational Study. PLoS One. 2015;10(6):e0130946.
119. Giugliano RP, Cannon CP, Blazing MA, Nicolau JC, Corbalán R, Špinar J, et al. Benefit of Adding Ezetimibe to Statin Therapy on Cardiovascular Outcomes and Safety in Patients With Versus Without Diabetes Mellitus: Results From IMPROVE-IT (Improved Reduction of Outcomes: Vytorin Efficacy International Trial). Circulation. 2018;137(15):1571-82.
120. Hassan S, Magny-Normilus C, Galusha D, Adams OP, Maharaj RG, Nazario CM, et al. Glycemic control and management of cardiovascular risk factors among adults with diabetes in the Eastern Caribbean Health Outcomes Research Network (ECHORN) Cohort Study. Prim Care Diabetes. 2022;16(1):107-15.
121. Han BH, Han K, Yoon KH, Kim MK, Lee SH. Impact of Mean and Variability of High-Density Lipoprotein-Cholesterol on the Risk of Myocardial Infarction, Stroke, and Mortality in the General Population. J Am Heart Assoc. 2020;9(7):e015493.
122. Boey E, Gay GM, Poh KK, Yeo TC, Tan HC, Lee CH. Visit-to-visit variability in LDL- and HDL-cholesterol is associated with adverse events after ST-segment elevation myocardial infarction: A 5-year follow-up study. Atherosclerosis. 2016;244:86-92.
123. Lee HJ, Lee SR, Choi EK, Han KD, Oh S. Low Lipid Levels and High Variability are Associated With the Risk of New-Onset Atrial Fibrillation. J Am Heart Assoc. 2019;8(23):e012771.
124. Clark D, 3rd, Nicholls SJ, St John J, Elshazly MB, Kapadia SR, Tuzcu EM, et al. Visit-to-visit cholesterol variability correlates with coronary atheroma progression and clinical outcomes. Eur Heart J. 2018;39(27):2551-8.
125. Park JB, Kim DH, Lee H, Hwang IC, Yoon YE, Park HE, et al. Mildly Abnormal Lipid Levels, but Not High Lipid Variability, Are Associated With Increased Risk of Myocardial Infarction and Stroke in "Statin-Naive" Young Population A Nationwide Cohort Study. Circ Res. 2020;126(7):824-35.

126. Ellison RC, Zhang Y, Qureshi MM, Knox S, Arnett DK, Province MA, et al. Lifestyle determinants of high-density lipoprotein cholesterol: the National Heart, Lung, and Blood Institute Family Heart Study. Am Heart J. 2004;147(3):529-35.
127. Sheng CS, Tian J, Miao Y, Cheng Y, Yang Y, Reaven PD, et al. Prognostic Significance of Long-term HbA1c Variability for All-Cause Mortality in the ACCORD Trial. Diabetes Care. 2020;43(6):1185-90.
128. Lee S, Liu T, Zhou J, Zhang Q, Wong WT, Tse G. Predictions of diabetes complications and mortality using hba1c variability: a 10-year observational cohort study. Acta Diabetologica. 2020.
129. Skrha J, Soupal J, Skrha J, Jr., Prazny M. Glucose variability, HbA1c and microvascular complications. Rev Endocr Metab Disord. 2016;17(1):103-10.
130. Chang CM, Hsieh CJ, Huang JC, Huang IC. Acute and chronic fluctuations in blood glucose levels can increase oxidative stress in type 2 diabetes mellitus. Acta Diabetol. 2012;49 Suppl 1:S171-7.
131. Roever L, Tse G, Biondi-Zoccai G. Variability of metabolic parameters and risk of heart failure: Can it be a marker of incident heart failure? Int J Cardiol. 2019;293:183-4.
132. Akrivos J, Ravona-Springer R, Schmeidler J, LeRoith D, Heymann A, Preiss R, et al. Glycemic control, inflammation, and cognitive function in older patients with type 2 diabetes. Int J Geriatr Psychiatry. 2015;30(10):1093-100.
133. Lee EY, Yang Y, Kim HS, Cho JH, Yoon KH, Chung WS, et al. Effect of visit-to-visit LDL-, HDL-, and non-HDL-cholesterol variability on mortality and cardiovascular outcomes after percutaneous coronary intervention. Atherosclerosis. 2018;279:1-9.
134. Lim LL, Tan AT, Moses K, Rajadhyaksha V, Chan SP. Place of sodium-glucose cotransporter-2 inhibitors in East Asian subjects with type 2 diabetes mellitus: Insights into the management of Asian phenotype. J Diabetes Complications. 2017;31(2):494-503.
135. Azoulay L, Suissa S. Sulfonylureas and the Risks of Cardiovascular Events and Death: A Methodological Meta-Regression Analysis of the Observational Studies. Diabetes Care. 2017;40(5):706-14.
136. Schramm TK, Gislason GH, Vaag A, Rasmussen JN, Folke F, Hansen ML, et al. Mortality and cardiovascular risk associated with different insulin secretagogues compared with metformin in type 2 diabetes, with or without a previous myocardial infarction: a nationwide study. Eur Heart J. 2011;32(15):1900-8.
137. Hung AM, Roumie CL, Greevy RA, Liu X, Grijalva CG, Murff HJ, et al. Kidney function decline in metformin versus sulfonylurea initiators: assessment of time-dependent contribution of weight, blood pressure, and glycemic control. Pharmacoepidemiol Drug Saf. 2013;22(6):623-31.
138. Kooy A, de Jager J, Lehert P, Bets D, Wulffele MG, Donker AJ, et al. Long-term effects of metformin on metabolism and microvascular and macrovascular disease in patients with type 2 diabetes mellitus. Arch Intern Med. 2009;169(6):616-25.
139. Chiasson JL, Josse RG, Gomis R, Hanefeld M, Karasik A, Laakso M, et al. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. Lancet. 2002;359(9323):2072-7.
140. Yi QY, Deng G, Chen N, Bai ZS, Yuan JS, Wu GH, et al. Metformin inhibits development of diabetic retinopathy through inducing alternative splicing of VEGF-A. Am J Transl Res. 2016;8(9):3947-54.
141. Zhang Z, Zhang X, Korantzopoulos P, Letsas KP, Tse G, Gong M, et al. Thiazolidinedione use and atrial fibrillation in diabetic patients: a meta-analysis. BMC Cardiovasc Disord. 2017;17(1):96.
142. Liu C, Liu R, Fu H, Li J, Wang X, Cheng L, et al. Pioglitazone attenuates atrial remodeling and vulnerability to atrial fibrillation in alloxan-induced diabetic rabbits. Cardiovasc Ther. 2017;35(5).
143. Xu D, Murakoshi N, Igarashi M, Hirayama A, Ito Y, Seo Y, et al. PPAR-gamma activator pioglitazone prevents age-related atrial fibrillation susceptibility by improving antioxidant capacity and reducing apoptosis in a rat model. J Cardiovasc Electrophysiol. 2012;23(2):209-17.

144. Shimano M, Tsuji Y, Inden Y, Kitamura K, Uchikawa T, Harata S, et al. Pioglitazone, a peroxisome proliferator-activated receptor-gamma activator, attenuates atrial fibrosis and atrial fibrillation promotion in rabbits with congestive heart failure. Heart Rhythm. 2008;5(3):451-9.
145. Korantzopoulos P, Kokkoris S, Kountouris E, Protopsaltis I, Siogas K, Melidonis A. Regression of paroxysmal atrial fibrillation associated with thiazolidinedione therapy. Int J Cardiol. 2008;125(3):e51-3.
146. Lee S, Zhou J, Chang C, Liu T, Chang D, Wong WT, et al. Comparative effects of sodium glucose cotransporter 2 (SGLT2) inhibitors and dipeptidyl peptidase-4 (DPP4) inhibitors on new-onset atrial fibrillation and stroke outcomes. medRxiv. 2021:2021.01.04.21249211.
147. Breiman L. Classification and regression trees: Routledge; 2017.
148. Ishwaran H, Kogalur UB, Gorodeski EZ, Minn AJ, Lauer MS. High-Dimensional Variable Selection for Survival Data. Journal of the American Statistical Association. 2010;105(489):205-17.
149. Hsich E, Gorodeski EZ, Blackstone EH, Ishwaran H, Lauer MS. Identifying important risk factors for survival in patient with systolic heart failure using random survival forests. Circulation: Cardiovascular Quality and Outcomes. 2011;4(1):39-45.
150. Miao F, Cai Y-P, Zhang Y-X, Li Y, Zhang Y-T. Risk prediction of one-year mortality in patients with cardiac arrhythmias using random survival forest. Computational and mathematical methods in medicine. 2015;2015.
151. Ward MM, Pajevic S, Dreyfuss J, Malley JD. Short‐term prediction of mortality in patients with systemic lupus erythematosus: Classification of outcomes using random forests. Arthritis Care & Research: Official Journal of the American College of Rheumatology. 2006;55(1):74-80.
152. Mamyrova G, O'Hanlon TP, Monroe JB, Carrick DM, Malley JD, Adams S, et al. Immunogenetic risk and protective factors for juvenile dermatomyositis in Caucasians. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 2006;54(12):3979-87.
153. Chen C, Zhou J, Yu H, Zhang Q, Gao L, Yin X, et al. Identification of important risk factors for all-cause mortality of acquired long QT syndrome patients using random survival forests and non-negative matrix factorization. Heart Rhythm. 2020.
154. Tse G, Zhou J, Lee S, Liu T, Bazoukis G, Mililis P, et al. Incorporating Latent Variables Using Nonnegative Matrix Factorization Improves Risk Stratification in Brugada Syndrome. J Am Heart Assoc. 2020:e012714.
155. Tse G, Lee S, Zhou J, Liu T, Wong ICK, Mak C, et al. Territory-Wide Chinese Cohort of Long QT Syndrome: Random Survival Forest and Cox Analyses. Front Cardiovasc Med. 2021;8:608592.
156. Hippisley-Cox J, Coupland C. Development and validation of QDiabetes-2018 risk prediction algorithm to estimate future risk of type 2 diabetes: cohort study. BMJ. 2017;359:j5019.
157. Palmer AJ, Roze S, Valentine WJ, Minshall ME, Foos V, Lurati FM, et al. The CORE Diabetes Model: Projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making. Curr Med Res Opin. 2004;20 Suppl 1:S5-26.
158. Shao H, Fonseca V, Stoecker C, Liu S, Shi L. Novel Risk Engine for Diabetes Progression and Mortality in USA: Building, Relating, Assessing, and Validating Outcomes (BRAVO). Pharmacoeconomics. 2018;36(9):1125-34.
159. Zhou H, Isaman DJ, Messinger S, Brown MB, Klein R, Brandle M, et al. A computer simulation model of diabetes progression, quality of life, and cost. Diabetes Care. 2005;28(12):2856-63.
160. Chan JCN, Cheung C-K, Cheung MYF, Swaminathan R, Critchley JAJH, Cockram CS. Abnormal Albuminuria as a Predictor of Mortality and Renal Impairment in Chinese Patients With NIDDM. Diabetes Care. 1995;18(7):1013-6.
161. Ma RCW, Chan JCN. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281(1):64-91.
162. Ha KH, Lee YH, Song SO, Lee JW, Kim DW, Cho KH, et al. Development and Validation of the Korean Diabetes Risk Score: A 10-Year National Cohort Study. Diabetes Metab J. 2018;42(5):402-14.

163. Wang A, Chen G, Su Z, Liu X, Liu X, Li H, et al. Risk scores for predicting incidence of type 2 diabetes in the Chinese population: the Kailuan prospective study. Sci Rep. 2016;6:26548.
164. Luo S, Han L, Zeng P, Chen F, Pan L, Wang S, et al. A risk assessment model for type 2 diabetes in Chinese. PLoS One. 2014;9(8):e104046.
165. Quan J, Pang D, Li TK, Choi CH, Siu SC, Tang SY, et al. Risk Prediction Scores for Mortality, Cerebrovascular, and Heart Disease Among Chinese People With Type 2 Diabetes. J Clin Endocrinol Metab. 2019;104(12):5823-30.
166. Scott ES, Januszewski AS, O'Connell R, Fulcher G, Scott R, Kesaniemi A, et al. Long-Term Glycemic Variability and Vascular Complications in Type 2 Diabetes: Post Hoc Analysis of the FIELD Study. J Clin Endocrinol Metab. 2020;105(10).
167. Wan EYF, Yu EYT, Chin WY, Ng FTY, Chia SMC, Wong ICK, et al. Age-specific associations of glycated haemoglobin variability with cardiovascular disease and mortality in patients with type 2 diabetes mellitus: A 10- year cohort study. Diabetes Obes Metab. 2020;22(8):1316-27.
168. Zhang L, Wang Y, Niu M, Wang C, Wang Z. Machine learning for characterizing risk of type 2 diabetes mellitus in a rural Chinese population: the Henan Rural Cohort Study. Sci Rep. 2020;10(1):4406.
169. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Medical Research Methodology. 2018;18(1):24.
170. Lau WC, Chan EW, Cheung CL, Sing CW, Man KK, Lip GY, et al. Association Between Dabigatran vs Warfarin and Risk of Osteoporotic Fractures Among Patients With Nonvalvular Atrial Fibrillation. JAMA. 2017;317(11):1151-8.
171. Tsushima E. Interpreting Results from Statistical Hypothesis Testing: Understanding the Appropriate P-value. Phys Ther Res. 2022;25(2):49-55.
172. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–58.
173. Ng AY. Feature selection, L₁ vs. L₂ regularization, and rotational invariance. Proceedings of the twenty-first international conference on Machine learning; Banff, Alberta, Canada: Association for Computing Machinery; 2004. p. 78.
174. Bergstra J, Bengio Y. Random Search for Hyper-Parameter Optimization. The Journal of Machine Learning Research. 2012;13:281-305.
175. Fox BL. Algorithm 647: Implementation and Relative Efficiency of Quasirandom Sequence Generators. ACM Trans Math Softw. 1986;12(4):362–76.
176. Schmidt-Hieber J. Nonparametric regression using deep neural networks with ReLU activation function. Annals of Statistics. 2017;48.
177. Griffith KN, Prentice JC, Mohr DC, Conlin PR. Predicting 5- and 10-Year Mortality Risk in Older Adults With Diabetes. Diabetes Care. 2020;43(8):1724-31.
178. Lowe G, Woodward M, Hillis G, Rumley A, Li Q, Harrap S, et al. Circulating inflammatory markers and the risk of vascular complications and mortality in people with type 2 diabetes and cardiovascular disease or risk factors: the ADVANCE study. Diabetes. 2014;63(3):1115-23.
179. Yi S-W, Yi J-J, Ohrr H. Total cholesterol and all-cause mortality by sex and age: a prospective cohort study among 12.8 million adults. Sci Rep. 2019;9(1):1596.
180. Sung KC, Huh JH, Ryu S, Lee JY, Scorletti E, Byrne CD, et al. Low Levels of Low-Density Lipoprotein Cholesterol and Mortality Outcomes in Non-Statin Users. J Clin Med. 2019;8(10).
181. Madsen CM, Varbo A, Nordestgaard BG. Extreme high high-density lipoprotein cholesterol is paradoxically associated with high mortality in men and women: two prospective cohort studies. Eur Heart J. 2017;38(32):2478-86.
182. Ma RC, Chan JC. Type 2 diabetes in East Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci. 2013;1281:64-91.

183. Lou M, Luo P, Tang R, Peng Y, Yu S, Huang W, et al. Relationship between neutrophil-lymphocyte ratio and insulin resistance in newly diagnosed type 2 diabetes mellitus patients. BMC Endocr Disord. 2015;15:9.
184. Azab B, Daoud J, Naeem FB, Nasr R, Ross J, Ghimire P, et al. Neutrophil-to-lymphocyte ratio as a predictor of worsening renal function in diabetic patients (3-year follow-up study). Ren Fail. 2012;34(5):571-6.
185. Tse G, Lai ETH, Tse V, Yeo JM. Molecular and Electrophysiological Mechanisms Underlying Cardiac Arrhythmogenesis in Diabetes Mellitus. J Diabetes Res. 2016;2016:2848759-.
186. Orozco-Beltran D, Navarro-Perez J, Cebrian-Cuenca AM, Alvarez-Guisasola F, Caride-Miana E, Mora G, et al. The influence of hemoglobin A1c levels on cardiovascular events and all-cause mortality in people with diabetes over 70 years of age. A prospective study. Prim Care Diabetes. 2020.
187. Jerez JM, Franco L, Alba E, Llombart-Cussac A, Lluch A, Ribelles N, et al. Improvement of breast cancer relapse prediction in high risk intervals using artificial neural networks. Breast Cancer Research and Treatment. 2005;94(3):265-72.
188. Faraggi D, Simon R. A neural network model for survival data. Statistics in Medicine. 1995;14(1):73-82.
189. Katzman J, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. Deep Survival: A Deep Cox Proportional Hazards Network. 2016.
190. Priori SG, Blomstrom-Lundqvist C, Mazzanti A, Blom N, Borggrefe M, Camm J, et al. 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death of the European Society of Cardiology (ESC). Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC). Eur Heart J. 2015;36(41):2793-867.
191. Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, et al. Worldwide epidemiology of atrial fibrillation: a Global Burden of Disease 2010 Study. Circulation. 2014;129(8):837-47.
192. Ball J, Carrington MJ, McMurray JJ, Stewart S. Atrial fibrillation: profile and burden of an evolving epidemic in the 21st century. Int J Cardiol. 2013;167(5):1807-24.
193. Marijon E, Le Heuzey JY, Connolly S, Yang S, Pogue J, Brueckmann M, et al. Causes of death and influencing factors in patients with atrial fibrillation: a competing-risk analysis from the randomized evaluation of long-term anticoagulant therapy study. Circulation. 2013;128(20):2192-201.
194. Benjamin EJ, Levy D, Vaziri SM, D'Agostino RB, Belanger AJ, Wolf PA. Independent risk factors for atrial fibrillation in a population-based cohort. The Framingham Heart Study. JAMA. 1994;271(11):840-4.
195. Kuo L, Chao TF, Liu CJ, Chen SJ, Tuan TC, Lin YJ, et al. Usefulness of the CHA2DS2-VASc Score to Predict the Risk of Sudden Cardiac Death and Ventricular Arrhythmias in Patients With Atrial Fibrillation. Am J Cardiol. 2018;122(12):2049-54.
196. Chao TF, Liu CJ, Tuan TC, Chen SJ, Chen TJ, Lip GYH, et al. Risk and Prediction of Sudden Cardiac Death and Ventricular Arrhythmias for Patients with Atrial Fibrillation - A Nationwide Cohort Study. Sci Rep. 2017;7:46445.
197. Aune D, Schlesinger S, Norat T, Riboli E. Diabetes mellitus and the risk of sudden cardiac death: A systematic review and meta-analysis of prospective studies. Nutr Metab Cardiovasc Dis. 2018;28(6):543-56.
198. Dimova R, Chakarova N, Grozeva G, Kirilov G, Tankova T. The relationship between glucose variability and insulin sensitivity and oxidative stress in subjects with prediabetes. Diabetes Res Clin Pract. 2019;158:107911.
199. Wentholt IM, Kulik W, Michels RP, Hoekstra JB, DeVries JH. Glucose fluctuations and activation of oxidative stress in patients with type 1 diabetes. Diabetologia. 2008;51(1):183-90.

200. Navab M, Reddy ST, Van Lenten BJ, Fogelman AM. HDL and cardiovascular disease: atherogenic and atheroprotective mechanisms. Nat Rev Cardiol. 2011;8(4):222-32.
201. Lee S, Liu T, Zhou J, Zhang Q, Wong WT, Tse G. Predictions of diabetes complications and mortality using hba1c variability: a 10-year observational cohort study. Acta Diabetol. 2020.
202. Lee S, Zhou J, Guo CL, Wong WT, Liu T, Wong ICK, et al. Predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death. Endocrinology, Diabetes & Metabolism. 2021;n/a(n/a):e00240.
203. Lee S, Zhou J, Wong WT, Liu T, Wu WKK, Wong ICK, et al. Glycemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning. BMC Endocr Disord. 2021.
204. Sertbas Y, Ozdemir A, Sertbas M, Dayan A, Sancak S, Uyan C. The Effect of Glucose Variability on QTc Duration and Dispersion in Patients with Type 2 Diabetes Mellitus. Pak J Med Sci. 2017;33(1):22-6.
205. Su JB, Yang XH, Zhang XL, Cai HL, Huang HY, Zhao LH, et al. The association of long-term glycaemic variability versus sustained chronic hyperglycaemia with heart rate-corrected QT interval in patients with type 2 diabetes. PLoS One. 2017;12(8):e0183055.
206. Murphy NP, Ford-Adams ME, Ong KK, Harris ND, Keane SM, Davies C, et al. Prolonged cardiac repolarisation during spontaneous nocturnal hypoglycaemia in children and adolescents with type 1 diabetes. Diabetologia. 2004;47(11):1940-7.
207. Takebayashi K, Aso Y, Matsutomo R, Wakabayashi S, Inukai T. Association between the corrected QT intervals and combined intimal-medial thickness of the carotid artery in patients with type 2 diabetes. Metabolism. 2004;53(9):1152-7.
208. Stern K, Cho YH, Benitez-Aguirre P, Jenkins AJ, McGill M, Mitchell P, et al. QT interval, corrected for heart rate, is associated with HbA1c concentration and autonomic function in diabetes. Diabet Med. 2016;33(10):1415-21.
209. Fitzpatrick C, Chatterjee S, Seidu S, Bodicoat DH, Ng GA, Davies MJ, et al. Association of hypoglycaemia and risk of cardiac arrhythmia in patients with diabetes mellitus: A systematic review and meta-analysis. Diabetes Obes Metab. 2018;20(9):2169-78.
210. Stahn A, Pistrosch F, Ganz X, Teige M, Koehler C, Bornstein S, et al. Relationship between hypoglycemic episodes and ventricular arrhythmias in patients with type 2 diabetes and cardiovascular diseases: silent hypoglycemias and silent arrhythmias. Diabetes Care. 2014;37(2):516-20.
211. Zunkler BJ. Human ether-a-go-go-related (HERG) gene and ATP-sensitive potassium channels as targets for adverse drug effects. Pharmacol Ther. 2006;112(1):12-37.
212. Saito S, Teshima Y, Fukui A, Kondo H, Nishio S, Nakagawa M, et al. Glucose fluctuations increase the incidence of atrial fibrillation in diabetic rats. Cardiovasc Res. 2014;104(1):5-14.
213. Li S, Zheng Z, Tang X, Zhong J, Liu X, Zhao Y, et al. Impact of HbA1c variability on subclinical left ventricular remodeling and dysfunction in patients with type 2 diabetes mellitus. Clin Chim Acta. 2020;502:159-66.
214. Paelestik KB, Jespersen NR, Jensen RV, Johnsen J, Botker HE, Kristiansen SB. Effects of hypoglycemia on myocardial susceptibility to ischemia-reperfusion injury and preconditioning in hearts from rats with and without type 2 diabetes. Cardiovasc Diabetol. 2017;16(1):148.
215. Yun JS, Park YM, Han K, Cha SA, Ahn YB, Ko SH. Severe hypoglycemia and the risk of cardiovascular disease and mortality in type 2 diabetes: a nationwide population-based cohort study. Cardiovasc Diabetol. 2019;18(1):103.
216. Kuroda M, Shinke T, Sakaguchi K, Otake H, Takaya T, Hirota Y, et al. Association between daily glucose fluctuation and coronary plaque properties in patients receiving adequate lipid-lowering therapy assessed by continuous glucose monitoring and optical coherence tomography. Cardiovasc Diabetol. 2015;14:78.

217. Sathyapalan T, Atkin SL, Kilpatrick ES. LDL cholesterol variability in patients with Type 2 diabetes taking atorvastatin and simvastatin: a comparison of two formulae for LDL-C estimation. Ann Clin Biochem. 2015;52(Pt 1):180-2.
218. Qamar A, Giugliano RP, Keech AC, Kuder JF, Murphy SA, Kurtz CE, et al. Interindividual Variation in Low-Density Lipoprotein Cholesterol Level Reduction With Evolocumab: An Analysis of FOURIER Trial Data. JAMA Cardiol. 2019;4(1):59-63.
219. Tanamas SK, Saulnier PJ, Hanson RL, Nelson RG, Hsueh WC, Sievers ML, et al. Serum lipids and mortality in an American Indian population: A longitudinal study. J Diabetes Complications. 2018;32(1):18-26.
220. Lu J, Ma X, Zhou J, Zhang L, Mo Y, Ying L, et al. Association of Time in Range, as Assessed by Continuous Glucose Monitoring, With Diabetic Retinopathy in Type 2 Diabetes. Diabetes Care. 2018;41(11):2370-6.
221. Adhyaru BB, Jacobson TA. Safety and efficacy of statin therapy. Nat Rev Cardiol. 2018;15(12):757-69.
222. Chang SH, Wu LS, Chiou MJ, Liu JR, Yu KH, Kuo CF, et al. Association of metformin with lower atrial fibrillation risk among patients with type 2 diabetes mellitus: a population-based dynamic cohort and in vitro studies. Cardiovasc Diabetol. 2014;13:123.
223. Shaw JE, Sicree RA, Zimmet PZ. Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Res Clin Pract. 2010;87(1):4-14.
224. Campbell PT, Newton CC, Patel AV, Jacobs EJ, Gapstur SM. Diabetes and cause-specific mortality in a prospective cohort of one million U.S. adults. Diabetes Care. 2012;35(9):1835-44.
225. Waks JW, Sitlani CM, Soliman EZ, Kabir M, Ghafoori E, Biggs ML, et al. Global Electric Heterogeneity Risk Score for Prediction of Sudden Cardiac Death in the General Population: The Atherosclerosis Risk in Communities (ARIC) and Cardiovascular Health (CHS) Studies. Circulation. 2016;133(23):2222-34.
226. Stevens RJ, Kothari V, Adler AI, Stratton IM, United Kingdom Prospective Diabetes Study G. The UKPDS risk engine: a model for the risk of coronary heart disease in Type II diabetes (UKPDS 56). Clin Sci (Lond). 2001;101(6):671-9.
227. Yang X, Li J, Hu D, Chen J, Li Y, Huang J, et al. Predicting the 10-Year Risks of Atherosclerotic Cardiovascular Disease in Chinese Population: The China-PAR Project (Prediction for ASCVD Risk in China). Circulation. 2016;134(19):1430-40.
228. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. JAMA. 2007;297(6):611-9.
229. Kontopantelis E, Springate DA, Reeves D, Ashcroft DM, Rutter MK, Buchan I, et al. Glucose, blood pressure and cholesterol levels and their relationships to clinical outcomes in type 2 diabetes: a retrospective cohort study. Diabetologia. 2015;58(3):505-18.
230. Aguilar D, Bozkurt B, Ramasubbu K, Deswal A. Relationship of hemoglobin A1C and mortality in heart failure patients with diabetes. J Am Coll Cardiol. 2009;54(5):422-8.
231. Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics. 2006;15(3):651-74.
232. Strobl C, Boulesteix A-L, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics. 2007;8(1):25.
233. Nichols GA, Joshua-Gotlib S, Parasuraman S. Glycemic control and risk of cardiovascular disease hospitalization and all-cause mortality. J Am Coll Cardiol. 2013;62(2):121-7.
234. Huang ES, Liu JY, Moffet HH, John PM, Karter AJ. Glycemic control, complications, and death in older diabetic patients: the diabetes and aging study. Diabetes Care. 2011;34(6):1329-36.
235. Lee Y, Park S, Lee S, Kim Y, Kang MW, Cho S, et al. Lipid profiles and risk of major adverse cardiovascular events in CKD and diabetes: A nationwide population-based study. PLoS One. 2020;15(4):e0231328.

236. Dworzynski P, Aasbrenn M, Rostgaard K, Melbye M, Gerds TA, Hjalgrim H, et al. Nationwide prediction of type 2 diabetes comorbidities. Sci Rep. 2020;10(1):1776.
237. Sakurai M, Saitoh S, Miura K, Nakagawa H, Ohnishi H, Akasaka H, et al. HbA1c and the risks for all-cause and cardiovascular mortality in the general Japanese population: NIPPON DATA90. Diabetes Care. 2013;36(11):3759-65.
238. An X, Yu D, Zhang R, Zhu J, Du R, Shi Y, et al. Insulin resistance predicts progression of de novo atherosclerotic plaques in patients with coronary heart disease: a one-year follow-up study. Cardiovasc Diabetol. 2012;11:71.
239. Iguchi T, Hasegawa T, Otsuka K, Matsumoto K, Yamazaki T, Nishimura S, et al. Insulin resistance is associated with coronary plaque vulnerability: insight from optical coherence tomography analysis. Eur Heart J Cardiovasc Imaging. 2014;15(3):284-91.
240. Erickson JR, Pereira L, Wang L, Han G, Ferguson A, Dao K, et al. Diabetic hyperglycaemia activates CaMKII and arrhythmias by O-linked glycosylation. Nature. 2013;502(7471):372-6.
241. Sifuentes-Franco S, Pacheco-Moises FP, Rodriguez-Carrizalez AD, Miranda-Diaz AG. The Role of Oxidative Stress, Mitochondrial Function, and Autophagy in Diabetic Polyneuropathy. J Diabetes Res. 2017;2017:1673081.
242. Reno CM, Daphna-Iken D, Chen YS, VanderWeele J, Jethi K, Fisher SJ. Severe hypoglycemia-induced lethal cardiac arrhythmias are mediated by sympathoadrenal activation. Diabetes. 2013;62(10):3570-81.
243. Landstedt-Hallin L, Englund A, Adamson U, Lins PE. Increased QT dispersion during hypoglycaemia in patients with type 2 diabetes mellitus. J Intern Med. 1999;246(3):299-307.
244. Marques JL, George E, Peacey SR, Harris ND, Macdonald IA, Cochrane T, et al. Altered ventricular repolarization during hypoglycaemia in patients with diabetes. Diabet Med. 1997;14(8):648-54.
245. Chow E, Bernjak A, Walkinshaw E, Lubina-Solomon A, Freeman J, Macdonald IA, et al. Cardiac Autonomic Regulation and Repolarization During Acute Experimental Hypoglycemia in Type 2 Diabetes. Diabetes. 2017;66(5):1322-33.
246. Investigators OT, Mellbin LG, Ryden L, Riddle MC, Probstfield J, Rosenstock J, et al. Does hypoglycaemia increase the risk of cardiovascular events? A report from the ORIGIN trial. Eur Heart J. 2013;34(40):3137-44.
247. Gordon DJ, Probstfield JL, Garrison RJ, Neaton JD, Castelli WP, Knoke JD, et al. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. Circulation. 1989;79(1):8-15.
248. Fanni G, Rosato R, Gentile L, Anselmino M, Frea S, Ponzo V, et al. Is HDL cholesterol protective in patients with type 2 diabetes? A retrospective population-based cohort study. J Transl Med. 2020;18(1):189.
249. Dai WS, LaPorte RE, Hom DL, Kuller LH, D'Antonio JA, Gutai JP, et al. Alcohol consumption and high density lipoprotein cholesterol concentration among alcoholics. Am J Epidemiol. 1985;122(4):620-7.
250. Oh IH, Hur JK, Ryoo JH, Jung JY, Park SK, Yang HJ, et al. Very high high-density lipoprotein cholesterol is associated with increased all-cause mortality in South Koreans. Atherosclerosis. 2019;283:43-51.
251. Silva AC, Santos L, Dinis-Oliveira RJ, Magalhaes T, Santos A. Sudden cardiac death in young adult. Cardiovasc Toxicol. 2014;14(4):379-86.
252. Lee S, Zhou J, Li KHC, Leung KSK, Lakhani I, Liu T, et al. Territory-wide Cohort Study of Brugada Syndrome in Hong Kong: Predictors of Long-Term Outcomes Using Random Survival Forests and Non-Negative Matrix Factorization. Open Heart. 2021.
253. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC, et al. Machine Learning to Predict the Risk of Incident Heart Failure Hospitalization Among Patients With Diabetes: The WATCH-DM Risk Score. Diabetes Care. 2019;42(12):2298-306.

254. Dietrich S, Floegel A, Troll M, Kuhn T, Rathmann W, Peters A, et al. Random Survival Forest in practice: a method for modelling complex metabolomics data in time to event analysis. Int J Epidemiol. 2016;45(5):1406-20.
255. Sloan RA, Haaland BA, Sawada SS, Lee IM, Sui X, Lee DC, et al. A Fit-Fat Index for Predicting Incident Diabetes in Apparently Healthy Men: A Prospective Cohort Study. PLoS One. 2016;11(6):e0157703.
256. Wright MN, Dankowski T, Ziegler A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. Statistics in Medicine. 2017;36(8):1272-84.
257. Bangalore S, Breazna A, DeMicco DA, Wun CC, Messerli FH. Visit-to-visit low-density lipoprotein cholesterol variability and risk of cardiovascular outcomes: insights from the TNT trial. J Am Coll Cardiol. 2015;65(15):1539-48.
258. Shao Q, Meng L, Lee S, Tse G, Gong M, Zhang Z, et al. Empagliflozin, a sodium glucose co-transporter-2 inhibitor, alleviates atrial remodeling and improves mitochondrial function in high-fat diet/streptozotocin-induced diabetic rats. Cardiovasc Diabetol. 2019;18(1):165.
259. Martinez M, Santamarina J, Pavesi A, Musso C, Umpierrez GE. Glycemic variability and cardiovascular disease in patients with type 2 diabetes. BMJ Open Diabetes Res Care. 2021;9(1).
260. Danne T, Nimri R, Battelino T, Bergenstal RM, Close KL, DeVries JH, et al. International Consensus on Use of Continuous Glucose Monitoring. Diabetes Care. 2017;40(12):1631-40.
261. Saisho Y. Glycemic variability and oxidative stress: a link between diabetes and cardiovascular disease? Int J Mol Sci. 2014;15(10):18381-406.
262. Yapanis M, James S, Craig ME, O'Neal D, Ekinci EI. Complications of Diabetes and Metrics of Glycemic Management Derived From Continuous Glucose Monitoring. J Clin Endocrinol Metab. 2022;107(6):e2221-e36.
263. Zhou JJ, Schwenke DC, Bahn G, Reaven P. Glycemic Variation and Cardiovascular Risk in the Veterans Affairs Diabetes Trial. Diabetes Care. 2018;41(10):2187-94.
264. Segar MW, Patel KV, Vaduganathan M, Caughey MC, Butler J, Fonarow GC, et al. Association of Long-term Change and Variability in Glycemia With Risk of Incident Heart Failure Among Patients With Type 2 Diabetes: A Secondary Analysis of the ACCORD Trial. Diabetes Care. 2020;43(8):1920-8.
265. Chan JSK, Lee YHA, Liu K, Hui JMH, Dee EC, Ng K, et al. HbA1c Variability and Cardiovascular Events in Patients with Prostate Cancer Receiving Androgen Deprivation Therapy. Eur Urol Open Sci. 2023;47:3-11.
266. Chang C, Zhou J, Chou OHI, Chan J, Leung KSK, Lee TTL, et al. Predictive value of neutrophil-to-lymphocyte ratio for atrial fibrillation and stroke in type 2 diabetes mellitus: The Hong Kong Diabetes Study. Endocrinol Diabetes Metab. 2023;6(1):e397.
267. Ying C, Liu T, Ling H, Cheng M, Zhou X, Wang S, et al. Glucose variability aggravates cardiac fibrosis by altering AKT signalling path. Diab Vasc Dis Res. 2017;14(4):327-35.
268. Manosroi W, Phimphilai M, Waisayanand N, Buranapin S, Deerochanawong C, Gunaparn S, et al. Glycated hemoglobin variability and the risk of cardiovascular events in patients with prediabetes and type 2 diabetes mellitus: A post-hoc analysis of a prospective and multicenter study. J Diabetes Investig. 2023;14(12):1391-400.
269. Hsu JC, Yang YY, Chuang SL, Yu CC, Lin LY. Higher long-term visit-to-visit glycemic variability predicts new-onset atrial fibrillation in patients with diabetes mellitus. Cardiovasc Diabetol. 2021;20(1):148.
270. Yuan T, Yang T, Chen H, Fu D, Hu Y, Wang J, et al. New insights into oxidative stress and inflammation during diabetes mellitus-accelerated atherosclerosis. Redox Biol. 2019;20:247-60.
271. Yi M, Cao Q, Tang WH, Liu Q, Ke X. Day-to-day fasting plasma glucose variability on the short-term prognosis of ST-segment elevation myocardial infarction: A retrospective cohort study. Clin Cardiol. 2022;45(12):1246-54.
272. Golden SH, Yajnik C, Phatak S, Hanson RL, Knowler WC. Racial/ethnic differences in the burden of type 2 diabetes over the life course: a focus on the USA and India. Diabetologia. 2019;62(10):1751-60.

273. Hsu WH, Lai CW, Chen SC, Chiou HC, Hsiao PJ, Shin SJ, et al. Greater Low-Density Lipoprotein Cholesterol Variability Increases the Risk of Cardiovascular Events in Patients with Type 2 Diabetes Mellitus. Endocr Pract. 2019;25(9):918-25.
274. Wan EYF, Yu EYT, Chin WY, Barrett JK, Mok AHY, Lau CST, et al. Greater variability in lipid measurements associated with cardiovascular disease and mortality: A 10-year diabetes cohort study. Diabetes Obes Metab. 2020;22(10):1777-88.
275. Chan JSK, Satti DI, Lee YHA, Hui JMH, Lee TTL, Chou OHI, et al. High visit-to-visit cholesterol variability predicts heart failure and adverse cardiovascular events: a population-based cohort study. Eur J Prev Cardiol. 2022;29(14):e323-e5.
276. Kwon S, Lee SR, Choi EK, Lee SH, Han KD, Lee SY, et al. Visit-to-visit variability of metabolic parameters and risk of heart failure: A nationwide population-based study. Int J Cardiol. 2019;293:153-8.
277. Hovingh GK, Ray KK, Boekholdt SM. Is Cholesteryl Ester Transfer Protein Inhibition an Effective Strategy to Reduce Cardiovascular Risk? CETP as a Target to Lower CVD Risk: Suspension of Disbelief? Circulation. 2015;132(5):433-40.
278. Park JB, Shin E, Lee JE, Lee SJ, Lee H, Choi SY, et al. Genetic Determinants of Visit-to-Visit Lipid Variability: Genome-Wide Association Study in Statin-Naïve Korean Population. Front Cardiovasc Med. 2022;9:811657.
279. Villines TC, Rodriguez Lozano P. Transitioning From Stenosis to Plaque Burden in the Cardiac CT Era: The Changing Risk Paradigm. J Am Coll Cardiol. 2020;76(24):2814-6.
280. Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. J Am Med Inform Assoc. 2019;26(12):1545-59.
281. Lee S, Zhou J, Wong WT, Liu T, Wu WKK, Wong ICK, et al. Glycemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning. BMC Endocr Disord. 2021;21(1):94.
282. Lee S, Jeevaratnam K, Liu T, Chang D, Chang C, Wong WT, et al. Risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: A population-based cohort study. Clin Cardiol. 2021;44(11):1602-12.
283. Lee S, Zhou J, Leung KSK, Wu WKK, Wong WT, Liu T, et al. Development of a predictive risk model for all-cause mortality in patients with diabetes in Hong Kong. BMJ Open Diabetes Res Care. 2021;9(1).
284. Benjamin KJM, Katipalli T, Paquola ACM. dRFEtools: dynamic recursive feature elimination for omics. Bioinformatics. 2023;39(8).
285. Gamal A, Barakat S, Rezk A. Letter to Editor (Response from author): Toward a universal electronic health record system. J Biomed Inform. 2021;117:103770.
286. Shamout F, Zhu T, Clifton DA. Machine Learning for Clinical Outcome Prediction. IEEE Rev Biomed Eng. 2021;14:116-26.
287. Lin SL. Intelligent Fault Diagnosis and Forecast of Time-Varying Bearing Based on Deep Learning VMD-DenseNet. Sensors (Basel). 2021;21(22).
288. Beaulieu-Jones B, Finlayson SG, Chivers C, Chen I, McDermott M, Kandola J, et al. Trends and Focus of Machine Learning Applications for Health Research. JAMA Netw Open. 2019;2(10):e1914051.
289. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to Machine Learning, Neural Networks, and Deep Learning. Transl Vis Sci Technol. 2020;9(2):14.
290. Wang G, Liu X, Ying Z, Yang G, Chen Z, Liu Z, et al. Optimized glycemic control of type 2 diabetes with reinforcement learning: a proof-of-concept trial. Nat Med. 2023;29(10):2633-42.
291. Manzini E, Vlacho B, Franch-Nadal J, Escudero J, Génova A, Reixach E, et al. Longitudinal deep learning clustering of Type 2 Diabetes Mellitus trajectories using routinely collected health records. J Biomed Inform. 2022;135:104218.

292. Deng Y, Lu L, Aponte L, Angelidi AM, Novak V, Karniadakis GE, et al. Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. NPJ Digit Med. 2021;4(1):109.

Appendices

Appendix 1. List of publications directly arising from the work described in this thesis

1. Lee, S., Liu, T., Chung, C.T., Reinhold, J., Vassiliou, V.S.*, Tse, G.* (2024) PowerAI-Diabetes: Review of glycemic and lipid variability to predict cardiovascular events in Chinese diabetic population. npj Metabolic Health and Disease. DOI: 10.1038/s44324-024-00012-7
2. Lee S., Liu T., Zhou J., Zhang Q., Wong W.T., Tse G. (2020) Predictions of diabetes complications and mortality using Hba1c variability: a 10-year observational cohort study. Acta Diabetologica. DOI: 10.1007/s00592-020-01605-6
3. Lee S., Zhou J., Wong W.T., Liu T., Wu W.K.K., Wong I.C.K., Zhang Q., Tse G. (2021) Glycemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning. BMC Endocrine Disorders. DOI: 10.1186/s12902-021-00751-4
4. Lee S.*, Zhou J.*, Leung K.S.K., Wu W.K.K., Wong W.T., Liu T., Wong I.C.K., Jeevaratnam K., Zhang Q., Tse G. (2021) Development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong. BMJ Open Diabetes Research & Care. DOI: 10.1136/bmjdrc-2020-001950. (joint first authorship)
5. Lee S., Jeevaratnam K., Liu T., Chang D., Chang C., Wong W.T., Wong I.C.K., Lip G., Tse G. (2021) Risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study. Clinical Cardiology. DOI: 10.1002/clc.23728
6. Lee S.*, Zhou J.*, Guo C.L., Wong W.T., Liu T., Wong I.C.K., Jeeevaratnam K., Zhang Q., Tse G. (2021) Predictive scores for identifying patients with type 2 diabetes mellitus

at risk of acute myocardial infarction and sudden cardiac death. Endocrinology, Diabetes & Metabolism. DOI: 10.1002/edm2.240. (joint first authorship)

Appendix 2. Statement regarding published material submitted in this thesis and nature of contribution

I confirm no part of the material in this PhD by Publication has previously been submitted by me for a degree in this or any other University.

The papers 1-6 in this thesis represents work undertaken by me in collaboration with coauthors. Confirmation of my contribution is provided by coauthors in each paper in Appendix 3.

1. Lee, S., Liu, T., Chung, C.T., Reinhold, J., Vassiliou, V.S.*, Tse, G.* (2024) PowerAI-Diabetes: Review of glycemic and lipid variability to predict cardiovascular events in Chinese diabetic population. npj Metabolic Health and Disease. DOI : 10.1038/s44324-024-00012-7

I have developed the research question, executed the relevant research, wrote the manuscript and led on the manuscript revision.

2. Lee S., Liu T., Zhou J., Zhang Q., Wong W.T., Tse G. (2020) Predictions of diabetes complications and mortality using Hba1c variability: a 10-year observational cohort study. Acta Diabetologica. DOI: 10.1007/s00592-020-01605-6

I have developed the research question, constructed the registry, and wrote the manuscript. I have also led on the data analysis and manuscript revision.

3. Lee S., Zhou J., Wong W.T., Liu T., Wu W.K.K., Wong I.C.K., Zhang Q., Tse G. (2021) Glycemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning. BMC Endocrine Disorders. DOI: 10.1186/s12902-021-00751-4

I have developed the research question, constructed the registry, wrote and revised the manuscript. I have collaborated with Zhou J. on the data analysis, in particular the machine learning analysis.

4. Lee S.*, Zhou J.*, Leung K.S.K., Wu W.K.K., Wong W.T., Liu T., Wong I.C.K., Jeevaratnam K., Zhang Q., Tse G. (2021) Development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong. BMJ Open Diabetes Research & Care. DOI: 10.1136/bmjdrc-2020-001950

I have developed the research question, constructed the registry, and wrote the manuscript. The data analysis was performed in collaboration with Zhou J. Co-authors provided suggestions and revisions.

5. Lee S., Jeevaratnam K., Liu T., Chang D., Chang C., Wong W.T., Wong I.C.K., Lip G., Tse G. (2021) Risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study. Clinical Cardiology. DOI: 10.1002/clc.23728

I have developed the research question, constructed the registry, and wrote the manuscript. Co-authors contributed with the extraction of raw data, and revision of the manuscript.

6. Lee S.*, Zhou J.*, Guo C.L., Wong W.T., Liu T., Wong I.C.K., Jeeevaratnam K., Zhang Q., Tse G. (2021) Predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death. Endocrinology, Diabetes & Metabolism. DOI: 10.1002/edm2.240

I have developed the research question, constructed the registry, and wrote the manuscript. I have collaborated with Zhou J. on the data analysis. Co-authors provided suggestions and revisions to the manuscript.

Appendix 3. Correspondence from coauthor confirming contribution

香港都會大學
HONG KONG METROPOLITAN UNIVERSITY

School of Nursing and Health Studies

PRIVATE AND CONFIDENTIAL

19th May 2023

Confirmation of Original Contributions to Publications for PhD

I am writing to confirm that Dr. Sharen Lee was responsible for leading and conducting the studies/publications below, including study design, processing of raw data to build the databases, database cleaning, leading statistical analysis, drafting and revision of the manuscripts. The machine learning models were developed by Dr. Jiandong Zhou, who therefore served as the joint lead authors for publications 3 and 6, but Sharen contributed to the overall direction of the studies.

1. Lee, S., Liu, T., Chung, C.T., Reinhold, J., Vassiliou, V.S.*, Tse, G.* (2024) PowerAI-Diabetes: Review of glycemic and lipid variability to predict cardiovascular events in Chinese diabetic population. npj Metabolic Health and Disease. DOI : 10.1038/s44324-024-00012-7

2. Lee S., Liu T., Zhou J., Zhang Q., Wong W.T., Tse G. (2020) Predictions of diabetes complications and mortality using Hba1c variability: a 10-year observational cohort study. Acta Diabetologica. DOI: 10.1007/s00592-020-01605-6

3. Lee S., Zhou J., Wong W.T., Liu T., Wu W.K.K., Wong I.C.K., Zhang Q., Tse G. (2021) Glycemic and lipid variability for predicting complications and mortality in diabetes mellitus using machine learning. BMC Endocrine Disorders. DOI: 10.1186/s12902-021-00751-4

4. Lee S.*, Zhou J.*, Leung K.S.K., Wu W.K.K., Wong W.T., Liu T., Wong I.C.K., Jeevaratnam K., Zhang Q., Tse G. (2021) Development of a predictive risk model for all-cause mortality in diabetic patients in Hong Kong. BMJ Open Diabetes Research & Care. DOI: 10.1136/bmjdrc-2020-001950

5. Lee S., Jeevaratnam K., Liu T., Chang D., Chang C., Wong W.T., Wong I.C.K., Lip G., Tse G. (2021) Risk stratification of cardiac arrhythmias and sudden cardiac death in type 2 diabetes mellitus patients receiving insulin therapy: a population-based cohort study. Clinical Cardiology. DOI: 10.1002/clc.23728

6. Lee S.*, Zhou J.*, Guo C.L., Wong W.T., Liu T., Wong I.C.K., Jeeevaratnam K., Zhang Q., Tse G. (2021) Predictive scores for identifying patients with type 2 diabetes mellitus at risk of acute myocardial infarction and sudden cardiac death. Endocrinology, Diabetes & Metabolism. DOI: 10.1002/edm2.240

Please do not hesitate to contact me if I can be of further assistance in supporting her PhD application.

Yours faithfully,

Prof. Gary Tse

BA (Hons.) (Cantab.), MBBS (Imperial), MA (Cantab.), (MA (Dubl.), MPH (Manc.), MHM (UNSW), MD (Cantab.), DM (Oxon.), PhD (Cantab.) RCPathME, MFPH (U.K.), FAcadMEd, FESC, FACC, FSCAI, FHRS, FRCP (Glasg., Edin., Lond., Ire.), FRCPCH, FFPH, FRCPath, SFHEA

Professor and Associate Dean (Innovations and Research)
School of Nursing and Health Studies, Hong Kong Metropolitan University, Hong Kong
Principal Investigator, Tianjin Key Laboratory of Ionic-Molecular Function of Cardiovascular Disease, Tianjin
Distinguished Professor, Chair of Cardiology, Department of Cardiology, The Second Hospital of Tianjin Medical University, Tianjin
People's Republic of China

Visiting Professor, Faculty of Health and Medical Sciences, University of Surrey, Guildford
Former Reader (Clinical) in Public Heath Medicine, Kent and Medway Medical School, University of Kent (permanent appointment until the retirement age) and Public Health Consultant, Medway Council
United Kingdom