Predicting the risk of new onset of type 2 diabetes following exposure of Statin within patient with coronary artery disease

Septi Melisa[1]; Christianus Heru Setiawan[2]; Muhammad Solihuddin Muhtar[1]; Phan Thanh-Phuc[1]; Nguyen Phung-Anh[3,4,5,6]; Jason C. Hsu[1,3,4,5*]

1. International Ph.D. program in Biotech and Healthcare Management, College of Management, Taipei Medical University, Taipei, Taiwan
2. Ph.D. Program, School of Pharmacy, Taipei Medical University, Taipei, Taiwan
3. Clinical Data Center, Office of Data Science, Taipei Medical University, Taipei, Taiwan
4. Research Center of Health Care Industry Data Science, College of Management, Taipei Medical University, Taipei, Taiwan
5. Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei Medical University, Taipei, Taiwan;
6. Graduate Institute of Data Science, College of Management, Taipei Medical University, Taipei, Taiwan;

*Corresponding author

Background

Statins are widely used to prevent the progression of atherosclerosis, particularly in patients with coronary artery disease. However, statin use has been associated with an increased risk of new-incident type 2 diabetes mellitus compared to non-statin use.[1-3] Statin use can impair insulin signaling, affect downstream metabolites, and disrupt adipocyte maturation and differentiation, potentially leading to the development of diabetes.[4, 5] This presents a trade-off, as the benefits of statins may greatly outweigh the risks, yet understanding their potential long-term side effects is crucial. Predicting the risk of type 2 diabetes can help the physician to do preventive treatment.[1] Therefore, this study aims to develop a prediction model for the risk of new-onset type 2 diabetes in patients with coronary artery disease using statins.

Method

This retrospective cohort study utilized data from the Taipei Medical University Clinical Research Database (TMUCRD), which has been mapped to the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM). The TMUCRD consist of clinical data from three hospitals affiliated with Taipei Medical University: Taipei Medical University Hospital (TMUH), Shuang Ho Hospital, and Wanfang Hospital, which all hospitals are located on the northern Taiwan.

We defined the target cohorts as patients who were prescribed rosuvastatin or atorvastatin, with the index date set as the initiation day of statin therapy. To be included in the study, participants needed at least 365 days of continuous observation prior to the index date. Inclusion criteria specified patients over 18 years of age with a prior diagnosis of coronary artery disease, excluding those with a pre-existing diagnosis of type 2 diabetes. The outcome of interest was the new onset of type 2 diabetes mellitus, defined as the occurrence of the condition at least 30 days after cohort initiation, with follow-up periods of 5 and 10 years.

We utilized several machine learning algorithms, including Lasso logistic regression and gradient boosting machine, to develop the prediction model. The models were trained on 75% of the data and tested on the remaining 25%. Model performance was evaluated using the area under the curve (AUC). The covariates included gender, age groups, condition group era, drug group era, observation days, CHADS2, CHADS2-VASC, and the Charlson Comorbidity Index. We employed Atlas version 2.13.0 for cohort construction and characterization, while patient-level prediction package was used to develop the prediction model.

Result

There were 11,084 patients using rosuvastatin and 20,573 using atorvastatin with a diagnosis of coronary artery disease prior to the index date. After 10 years of exposure, 1,474 patients (13.6%) in the rosuvastatin group and 3,438 patients (16.2%) in the atorvastatin group developed new-onset diabetes. For baseline characteristics, rosuvastatin was predominantly used by patients aged 60-64 (16.53%), while atorvastatin was most commonly used by patients aged 55-59 (15.24%). Males constituted the majority in both exposure cohorts (see Table 1).

Table 1. Baseline Characteristics of Rosuvastatin and Atorvastatin group

Characteristics	Rosuvastatin (N = 11,084)	Atorvastatin (N = 20,573)
Age Group		
30 - 34	65 (0.59%)	176 (0.81%)
35 - 39	270 (2.44%)	511 (2.34%)
40 - 44	512 (4.62%)	1,031 (4.72%)
45 - 49	809 (7.30%)	1,706 (7.80%)
50 - 54	1,324 (11.95%)	2,656 (12.15%)
55 - 59	1,727 (15.58%)	3,332 (15.24%)
60 - 64	1,832 (16.53%)	3,307 (15.13%)
65 - 69	1,483 (13.38%)	2,827 (12.93%)
70 - 74	1,126 (10.16%)	2,083 (9.53%)
75 - 79	762 (6.87%)	1,680 (7.69%)
80 - 84	630 (5.68%)	1,314 (6.01%)
85 - 89	354 (3.19%)	800 (3.66%)
90 - 94	127 (1.15%)	298 (1.36%)
Gender		
Male	6,206 (55.99%)	12,397 (56.72%)
Female	4,878 (44.01%)	9,461 (43.28%)
Charlson Index – Romano Adaptation	1.61 (N = 8,020)	1.57 (N = 15,830)
Chads2	1.17 (N = 11,084)	1.21 (21,860)
Chads2Vasc	2.27 (N = 11,084)	2.29 (21,860)

The most effective model for predicting new-onset type 2 diabetes over a 10-year risk period post-cohort initiation achieved an AUC of 80.7% for the rosuvastatin group using logistic

regression, and an AUC of 79.6% for the atorvastatin group with the same model (see Figures 1 and 2). Table 2 provides a comprehensive summary of the performance metrics for all models.

Table 2. The model performance for predicting new onset of type 2 diabetes

Target	TAR	Model	Incidence Rate	Training AUC	Testing AUC	Sensitivity	Specificity	PPV	NPV
Rosuvastatin	5 years	LR	11.8%	81.1%	76.0%	68.7%	66.9%	21.7%	94.1%
		XGBoost		87.4%	75.7%	69%	68.1%	22.7%	94.2%
	10 years	LR	13.618%	**81.4%**	**80.7%**	**75.1%**	**73.2%**	**30.6%**	**94.9%**
		XGBoost		**81.8%**	**80.0%**	**72.6%**	**71%**	**28.3%**	**94.3%**
Atorvastatin	5 years	LR	13.399%	79%	77.6%	69%	68.1%	25%	93.4%
		XGBoost		86.2%	78.8%	70.5%	69.5%	26.3%	93.8%
	10 years	LR	16.274%	81.6%	79.6%	71.6%	71%	31.5%	92.8%
		XGBoost		79.9%	78.8%	70.2%	69.3%	30.7%	92.3%

TAR: time at risk, LR: Logistic Regression, XGBoost: Extra Gradient Boosting, PPV: Positive Predictive Value, NPV: Negative Predictive Value

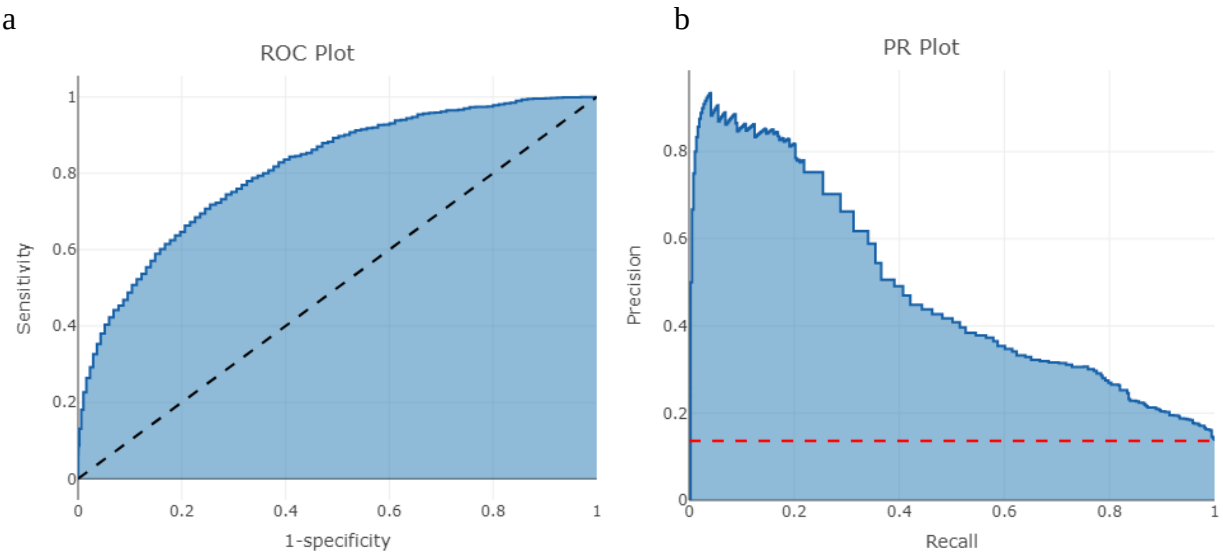

Figure 1. Area under curve (AUC) for predicting the risk of diabetes 10 years after rosuvastatin exposure, a. Testing AUC, b. Precision curve.

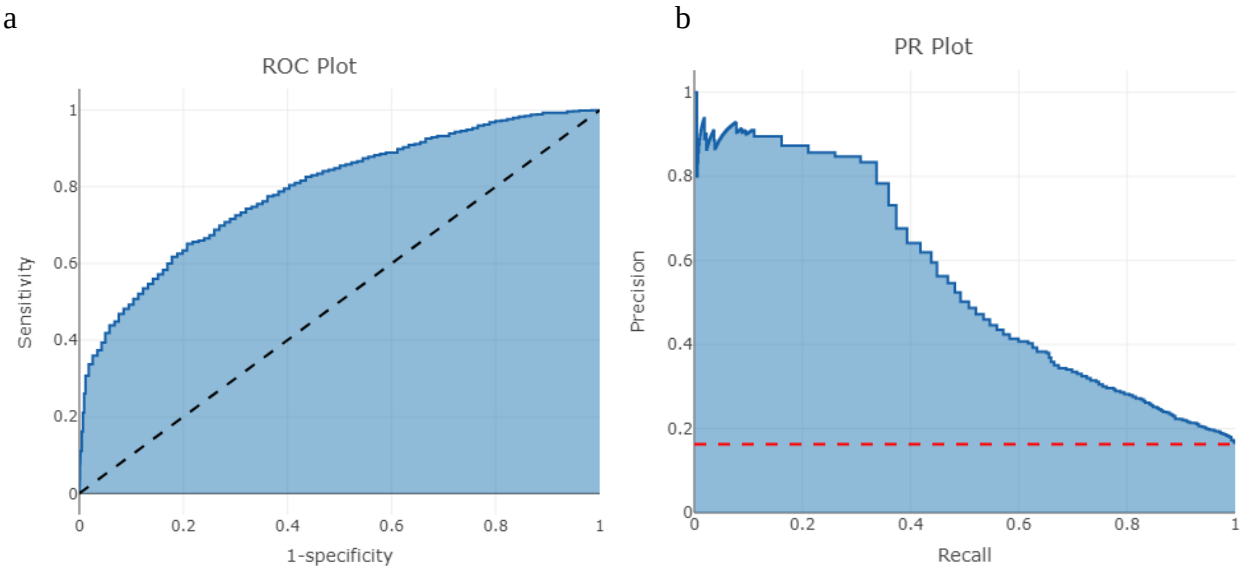

Figure 2. Area under curve (AUC) for predicting the risk of diabetes 10 years after atorvastatin exposure, a. Testing AUC, b. Precision curve.

Conclusion

We developed a prediction model to assess the risk of new-onset diabetes in patients with coronary artery disease undergoing treatment with statins such as rosuvastatin and atorvastatin. The model demonstrated good performance for each statin group. Predicting the new onset of diabetes enables physicians to advise high-risk patients on lifestyle modifications, regular diabetes screening, and early treatment initiation. Further studies involving larger sample size with diverse population characteristics are needed to improve and validate the model performance. Through this collaboration showcase, we would like invite data partners within the OHDSI community to collaborate with us in validating these findings and enhancing the robustness of our study.

Reference

1. Laakso M, Fernandes Silva L. Statins and risk of type 2 diabetes: mechanism and clinical implications. Front Endocrinol (Lausanne). 2023;14:1239335.
2. Reith C, Preiss D, Blackwell L, Emberson J, Spata E, Davies K, et al. Effects of statin therapy on diagnoses of new-onset diabetes and worsening glycaemia in large-scale randomised blinded statin trials: an individual participant data meta-analysis. The Lancet Diabetes & Endocrinology. 2024;12(5):306-19.
3. Mansi IA, Sumithran P, Kinaan M. Risk of diabetes with statins. BMJ. 2023;381:e071727.
4. Paseban M, Butler AE, Sahebkar A. Mechanisms of statin-induced new-onset diabetes. Journal of Cellular Physiology. 2019;234(8):12551-61.
5. Crandall JP, Mather K, Rajpathak SN, Goldberg RB, Watson K, Foo S, et al. Statin use and risk of developing diabetes: results from the Diabetes Prevention Program. BMJ Open Diabetes Research & Care. 2017;5(1):e000438.