Predicting Diabetic Eye Disease with AI: Trajectory Embedding and Causal Exploration of Statin Effects

Liyuan Zhu
NIHR Newcastle PSRC

Authors: Liyuan Zhu, John Casement, Konstantin Shestopaloff, Dexter Canoy, Michael R Barnes and Nick J Reynolds

Background: Diabetes patients often experience Multiple Long-Term Conditions (MLTCs), with cardiovascular disease as one of the most common. Statins are widely prescribed for cardiovascular risk reduction, but their longitudinal impact on diabetes and its complications, such as retinopathy remains unclear. Additionally, the impact of socioeconomic factors and gender inequalities on the burden and progression of MLTCs is not well understood, potentially aggravating inequities in patient safety.

Objectives: This study aims to: (1) Map and model the longitudinal MLTC trajectories in statin-treated diabetes patients. (2) Investigate the predictive value of MLTC trajectories for diabetic eye disease using Artificial Intelligence (AI). (3) Evaluate the causal effect of early statin initiation on diabetic eye disease. (4) Explore how socioeconomic and gender disparities influence the trajectory of MLTCs and potential patient outcomes.

Methods: Our study includes 13,391 participants in the UK Biobank with diabetes and extracted demographics data at baseline (2006-2010) and longitudinal medical histories (from their oldest EHR until 2019). Townsend Deprivation Index (TDI) is considered a socioeconomic indicator. Statin initiation, diabetes onset and MLTC trajectories were encoded using time-sequential embeddings. We trained multiple machine learning models to predict future diabetic eye disease. Causal inference was conducted to estimate the effect of statin initiation timing. Additionally, our research integrated Patient and Public Involvement and Engagement (PPIE) through regular weekly meetings, enabling patients to contribute insights and shape the research.

Results: Mapping the first 15 MLTCs, LightGBM achieved an AUC of 0.87 and F1-score of 0.80 in predicting diabetic eye disease, significantly improved predictive performance compared with the baseline model (AUC: 0.74, F1-score: 0.67). Patients with type 2 diabetes exhibited higher TDI scores than type 1 or unspecified diabetes patients, suggesting greater socioeconomic deprivation. Gender analyses revealed significant inequalities: female patients experienced both higher incidence rates and earlier onset of most MLTCs than males. Furthermore, causal analysis estimated Average Treatment Effect (ATE) using Propensity Score Matching (PSM) and Inverse Probability of Treatment Weighting (IPTW), showed early statin initiation was associated with a reduced risk of retinopathy (ATE: -0.63 via PSM; -0.21 via IPTW).

Conclusion: MLTC trajectories provide powerful, interpretable inputs for AI prediction of diabetic complications. The protective role of early statin use in retinopathy suggests a potential beneficial window for intervention, offering actionable insights into addressing patient safety inequities. Continuous PPIE has enriched our research perspective and ensured relevance to patient priorities.

Predicting Diabetic Eye Disease with AI: Trajectory Embedding and Causal Exploration of Statin Effects

Liyuan Zhu,[1] John Casement,[2] Konstantin Shestopaloff,[1] Dexter Canoy,[1] Sohan Seth,[3] Mike R. Barnes[4] and Nick J. Reynolds[5]

1. Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
2. Bioinformatics Support Unit, Newcastle University, Newcastle upon Tyne, UK
3. School of Informatics, University of Edinburgh, Edinburgh, UK
4. Centre for Translational Bioinformatics, William Harvey Research Institute, Queen Mary University of London, London, UK
5. Translational and Clinical Research Institute and NIHR Biomedical Research Centre, Newcastle University, Newcastle upon Tyne, UK

NIHR SafetyNet Symposium 2025

psrc-network.nihr.ac.uk

BACKGROUND

Diabetes patients often develop **Multiple Long-Term Conditions (MLTCs),** with cardiovascular disease being one of the most common. **Statins** are widely prescribed for cardiovascular risk reduction, but their longitudinal impact on diabetes and diabetic complications such as retinopathy remains unclear.

AIM

This study aims to predict **diabetic eye disease (DED)** using **AI** models with the **comorbidity trajectory** and explore the potential role of statins.

METHODS

(a) *Comorbidity Trajectory Analysis in Statin-Treated Diabetes Patients.* We analysed 13,391 patients within the UK Biobank, focusing on the evolution of MLTCs. By aligning disease onset sequences with statin initiation, we visualised comorbidity trajectories.
(b) *Predict Diabetic Eye Disease with Trajectories.* We encoded trajectories and trained machine learning models and compared the performance of the baseline model with the trajectory-incorporated models.
(c) *Ongoing Causal Inference on Statin Timing and Retinopathy Risk.* We estimated the average treatment effect of statin use.

RESULTS

Results 1.
Tables 2 and 3 show that most patients had diabetes before statin initiation. There is no evidence that statin use increases co-medication in patients with new-onset diabetes. Those who experience a worsening of diabetes (more antidiabetics prescribed) tend to be patients with pre-existing diabetes.
Results 2.
Figure 1 Top left: the timing of LTCs and complications in relation to the diagnosis of diabetes. **Middle and right:** the prescribing of statins largely coincided with the onset of diabetes although was slightly later than the first diagnosis of diabetes. **Bottom:** 47.5% of patients developed diabetic eye disease. Although males were over-represented (M:F = 62%:38%), females were more likely to have earlier co-LTCs, whether having diabetes or using statins at baseline.
Results 3.
We used machine learning to model the risk of developing DED. Incorporating LTCs and onset age, LightGBM achieved an AUC of 0.82 (Table 4), substantially outperforming the baseline model (AUC: 0.74) that relied solely on static features (Figure 3).
Results 4.
Model interpretation showed early statin initiation was associated with a *reduced* risk of retinopathy, by providing the contributions of statin timing to prediction (Figure 4), showing **earlier statin position** relates to a lower number of DED patients (Figure 5), and suggesting **early users** have a **63.4% lower risk** by PSM and a **21.0% lower risk** by IPTW (Figure 6).

CONCLUSION

Comorbidities, along with age of onset, provide powerful, interpretable inputs for AI prediction of diabetic complications. The protective role of early statin use in retinopathy suggests a potential beneficial **window for intervention**.

(a) Comorbidity Trajectory Analysis

	All cohort (n=13,391)	Statins first (3,774, 28%)	Diabetes first (8,933, 67%)	Same day (684, 5%)
Male	8,341 (62%)	2,455 (65%)	5,462 (61%)	424 (62%)
Female	5,050 (38%)	1,319 (35%)	3,471 (39%)	260 (38%)
Average age at first statin	58 (SD=7.6)	57 (SD=7.1)	58 (SD=7.8)	58 (SD=7.5)
Average age at first diabetes	55 (SD=10.7)	62 (SD=7.2)	52 (SD=10.9)	58 (SD=7.5)
Average total LTCs	10 (SD=4.9)	11 (SD=4.8)	10 (SD=4.9)	8 (SD=4.2)
Average LTCs before statins	5 (SD=3.7)	4 (SD=3.2)	6 (SD=3.7)	3 (SD=3.4)

Table 1. Descriptive statistics by event order, sex and LTC

All cohort n=13,391	Statins first n=3,774 (28%)	Diabetes first n=8,933 (67%)	Same day n=684 (5%)
Type 1 diabetes: 571 (4.3%)	8 (0.2%)	557 (6.2%)	6 (0.9%)
Type 2 diabetes: 12,159 (90.8%)	3659 (97.0%)	7855 (87.9%)	645 (94.3%)
Unspecified or Rare Diabetes: 661 (4.9%)	107 (2.8%)	521 (5.8%)	33 (4.8%)

Table 2. Descriptive statistics by event order and diabetes type

	All cohort n=13,391	Statins first n=3,774	Diabetes first n=8,933	Same day n=684
Mean number of antidiabetics before statins	0.86	0	1.29	0
Mean total antidiabetics	3.50	2.19	4.12	2.59

Table 3. Descriptive statistics by event order and the number of antidiabetics

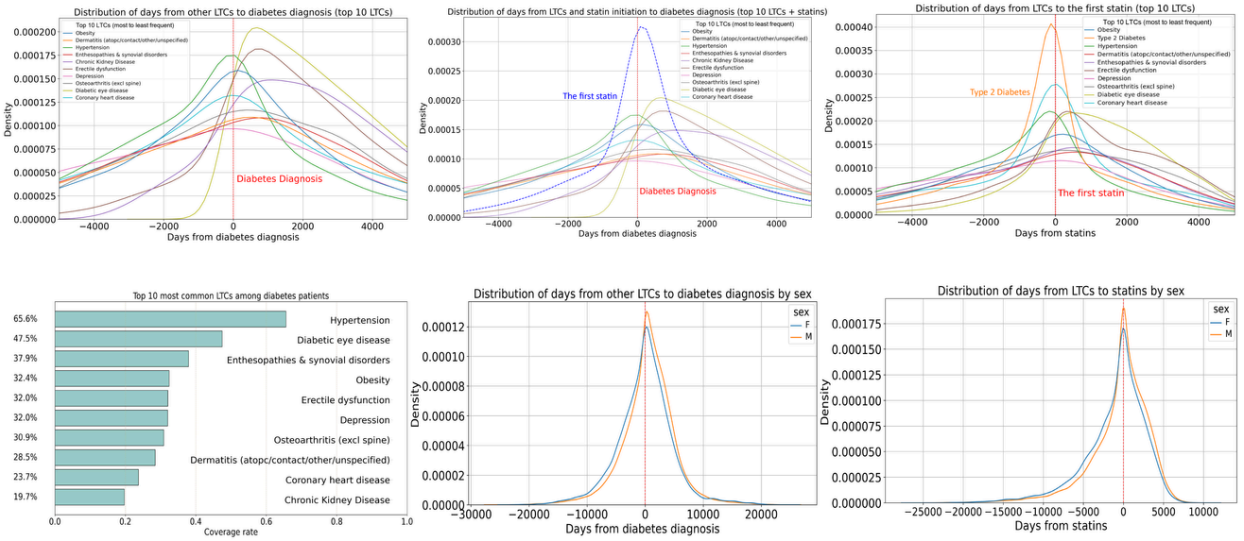

Figure 1. MLTC trajectory, frequency, and onset time by sex

(b) Prediction of Diabetic Eye Disease

		Precision	Recall	F1-score	Accuracy	AUC
XGBoost with all baseline features	0	0.64	0.67	0.65	0.63	0.68
	1	0.61	0.58	0.60		
XGBoost with all baseline features + SMOTE	0	0.65	0.66	0.66	0.65	0.71
	1	0.65	0.64	0.65		
LightGBM with all baseline features (baseline model)	0	0.70	0.66	0.68	0.67	0.74
	1	0.64	0.68	0.66		
LightGBM with all baseline features + SMOTE	0	0.70	0.70	0.70	0.69	0.78
	1	0.69	0.69	0.69		
Baseline model + LTCs	0	0.73	0.76	0.74	0.73	0.80
	1	0.72	0.69	0.70		
Baseline model + LTCs and onset age	**0**	**0.75**	**0.77**	**0.76**	**0.75**	**0.82**
	1	**0.74**	**0.72**	**0.73**		

Table 4. Model comparison

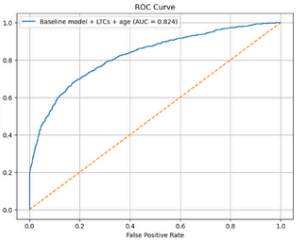

Figure 2. ROC plot for the best model

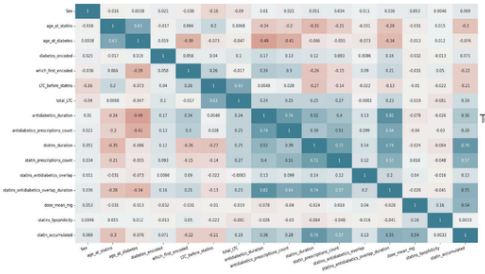

Figure 3. Baseline features and correlation matrix

(c) Causal Inference on Statin Timing and Retinopathy Risk

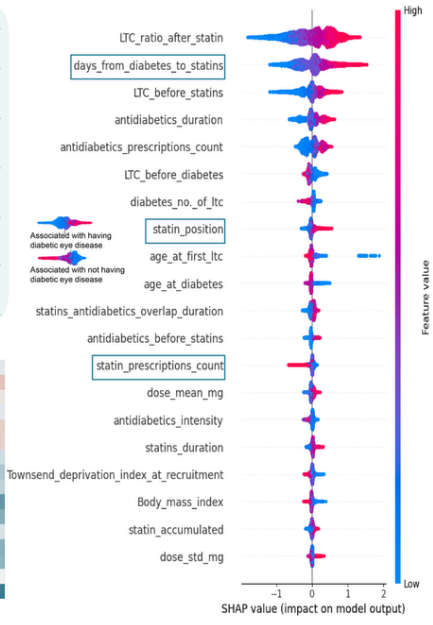

Figure 4. SHAP plot for model interpretation

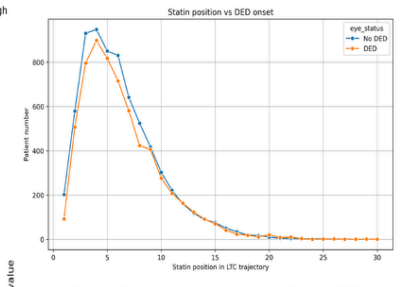

Figure 5. The relation between statin position and DED

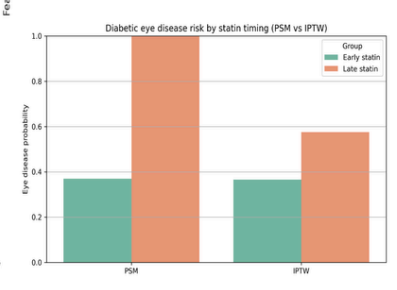

Figure 6. Causal Inference using Propensity Score Matching and Inverse Probability of Treatment Weighting

REFERENCES

[1] Reith, Christina 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, Volume 12, Issue 5, 306 - 319 (2024).
[2] Eswari, M.S., Balamurali, S. & Ramkumar, K. An Intelligent System to Predict Diabetic Retinal Diseases Based on Diabetic Attributes. SN COMPUT. SCI. 5, 755 (2024).
[3] Dai, L., Sheng, B., Chen, T. et al. A deep learning system for predicting time to progression of diabetic retinopathy. Nat Med 30, 584–594 (2024).
[4] Nderitu, P., Nunez do Rio, J.M., Webster, L. et al. Predicting 1, 2 and 3 year emergent referable diabetic retinopathy and maculopathy using deep learning. Commun Med 4, 167 (2024).
[5] Al-Janabi, A., Lightman, S. & Tomkins-Netzer, O. Statins in retinal disease. Eye 32, 981–991 (2018).
[6] Tomkins-Netzer, O., Niederer, R., Lightman, S. The role of statins in diabetic retinopathy. Trends in Cardiovascular Medicine 34, 128-135 (2024)

Correspondence to:
Liyuan Zhu: Newcastle Patient Safety Research Collaboration, Newcastle University, Newcastle upon Tyne, NE1 7RU, United Kingdom
l.zhu20@newcastle.ac.uk

This study is funded by the National Institute for Health and Care Research (NIHR) Newcastle Patient Safety Research Collaboration. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.