

## Development and Validation of a Diabetic Retinopathy Risk Stratification Algorithm

Dariusz Tarasewicz, Andrew J. Karter, Noel Pimentel, Howard H. Moffet, Khanh K. Thai, David Schlessinger, Oleg Sofrygin, and Ronald B. Melles

*Diabetes Care* 2023;46(5):1068–1075 | <https://doi.org/10.2337/dc22-1168>

### Development and Validation of a Diabetic Retinopathy Risk Stratification Algorithm



700 million people are projected to have diabetes by 2045.



Effective diabetic retinopathy screening can reduce risk of blindness by 90%.



Access to health services limits screening success and efficiency.



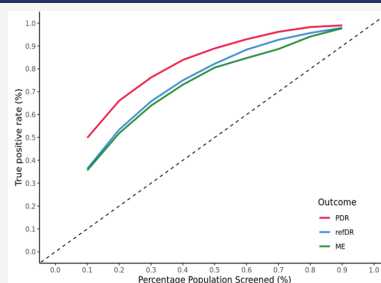
Can clinical risk stratification improve diabetic retinopathy screening?



276,794 Kaiser Permanente Northern California patients with diabetes included in retrospective study.



Risk stratification models were tested and validated for vision threatening diagnoses.



#### CONCLUSIONS:

At all screening capacities, model-based approach outperforms random screening (diagonal line) for all three outcome measures. Abbreviations: PDR (proliferative diabetic retinopathy), refDR (referable retinopathy), ME (diabetic macular edema).

### ARTICLE HIGHLIGHTS

- Diabetic retinopathy (DR) is a leading cause of blindness. Though timely screening and intervention can prevent 90% of DR-related blindness, barriers exist.
- This study used longitudinal clinical data from the electronic health record to create risk stratification models for the development of DR.
- Machine learning algorithms and logistic regression were trained using a primary outcome of proliferative DR and secondary outcomes of referable retinopathy and diabetic macular edema. A nine-covariate simple model had high predictive power for all three outcomes.
- Risk stratification tools greatly improve the efficiency of DR screening.



# Development and Validation of a Diabetic Retinopathy Risk Stratification Algorithm

*Diabetes Care* 2023;46:1068–1075 | <https://doi.org/10.2337/dc22-1168>

Dariusz Tarasewicz,<sup>1</sup> Andrew J. Karter,<sup>2</sup> Noel Pimentel,<sup>2</sup> Howard H. Moffet,<sup>2</sup> Khanh K. Thai,<sup>2</sup> David Schlessinger,<sup>2</sup> Oleg Sofrygin,<sup>2</sup> and Ronald B. Melles<sup>1</sup>

## OBJECTIVE

Although diabetic retinopathy is a leading cause of blindness worldwide, diabetes-related blindness can be prevented through effective screening, detection, and treatment of disease. The study goal was to develop risk stratification algorithms for the onset of retinal complications of diabetes, including proliferative diabetic retinopathy, referable retinopathy, and macular edema.

## RESEARCH DESIGN AND METHODS

Retrospective cohort analysis of patients from the Kaiser Permanente Northern California Diabetes Registry who had no evidence of diabetic retinopathy at a baseline diabetic retinopathy screening during 2008–2020 was performed. Machine learning and logistic regression prediction models for onset of proliferative diabetic retinopathy, diabetic macular edema, and referable retinopathy detected through routine screening were trained and internally validated. Model performance was assessed using area under the curve (AUC) metrics.

## RESULTS

The study cohort ( $N = 276,794$ ) was 51.9% male and 42.1% White. Mean ( $\pm$ SD) age at baseline was 60.0 ( $\pm$ 13.1) years. A machine learning XGBoost algorithm was effective in identifying patients who developed proliferative diabetic retinopathy (AUC 0.86; 95% CI, 0.86–0.87), diabetic macular edema (AUC 0.76; 95% CI, 0.75–0.77), and referable retinopathy (AUC 0.78; 95% CI, 0.78–0.79). Similar results were found using a simpler nine-covariate logistic regression model: proliferative diabetic retinopathy (AUC 0.82; 95% CI, 0.80–0.83), diabetic macular edema (AUC 0.73; 95% CI, 0.72–0.74), and referable retinopathy (AUC 0.75; 95% CI, 0.75–0.76).

## CONCLUSIONS

Relatively simple logistic regression models using nine readily available clinical variables can be used to rank order patients for onset of diabetic eye disease and thereby more efficiently prioritize and target screening for at risk patients.

Diabetes has been described as a global pandemic (1), with over 700 million people projected to have diabetes by 2045 (2). The societal cost of diabetes is already significant, with diabetes care representing approximately 25% of current Medicare expenditures (3). One of the most feared complications of diabetes is vision loss due to diabetic retinopathy (DR), which occurs in approximately one-third of patients with DM.

The International Clinical Diabetic Retinopathy Disease Severity Scale separates DR into five categories based on retina fundus signs: no retinopathy, or mild,

<sup>1</sup>Department of Ophthalmology, The Permanente Medical Group, Oakland, CA

<sup>2</sup>Division of Research, Kaiser Permanente, Oakland, CA

Corresponding author: Dariusz Tarasewicz, [dariusz.tarasewicz@kp.org](mailto:dariusz.tarasewicz@kp.org)

Received 14 June 2022 and accepted 23 February 2023

This article contains supplementary material online at <https://doi.org/10.2337/figshare.22155383>.

© 2023 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. More information is available at <https://www.diabetesjournals.org/journals/pages/license>.

moderate, severe, and proliferative retinopathy (4). Patients with moderate or higher levels of retinopathy, also called referable retinopathy (refDR), are at increased risk of vision loss and usually require additional ancillary testing or close follow-up by an ophthalmologist. Diabetic macular edema (DME) or proliferative DR (PDR) are manifestations of sight-threatening retinopathy that can cause significant vision loss if not identified and treated promptly.

Although 90% of blindness from DR can be prevented by early detection and treatment (5), screening for DR in the ever-increasing population with diabetes represents a significant effort for health care delivery systems (6,7). Fundus photography has become the standard for accurate and rapid identification of DR (8–11) but in most urban centers, fewer than 50% of the population with diabetes receives DR screening (12). Failure to identify DR in a timely manner leads to an increased risk of blindness and a reduced quality of life.

Health care systems screen for DR through direct clinical exams or nonmydriatic fundus photography. One strategy to improve DR screening involves the use of artificial intelligence to read and interpret digital fundus photos (13,14). An alternate approach to improve screening efficiency is to prioritize screening among patients who are at high risk based on systemic, nonocular risk factors. At any given time, over 85% of patients screened by Healthcare Effectiveness Data and Information Set guidelines show no retinopathy or mild retinopathy and require no treatment. Prioritizing those at highest risk for visual complications based on a combination of clinical characteristics has the potential to create greater efficiencies in the screening process.

By leveraging electronic medical records (EMR) in integrated health care delivery systems, population-based studies of DR can now be undertaken on much larger cohorts compared with the smaller studies performed in the 1990s (15–17). With the more complete and wider range of clinical data captured in the EMR, associations between nonocular clinical risk factors and DR onset might be better understood.

In the current study, machine learning (ML) and logistic regression (LR) predictive models for DR were developed and validated with the goal of facilitating risk stratification of a large population with

diabetes from Kaiser Permanente Northern California (KPNC).

## RESEARCH DESIGN AND METHODS

### Data Collection

This was a retrospective observational cohort study of patients with diabetes in KPNC, one of the largest integrated health care delivery systems in the U.S. Inclusion criteria were patients in the KPNC Diabetes Registry (18) aged 18 years or older with  $\geq 1$  DR screenings between 2 January 2008 and 31 December 2020 with no evidence of retinopathy. A patient's earliest negative retinopathy screening was the baseline for follow-up. Each study participant was also required to have continuous KPNC membership for 12 months prior to baseline. The study cohort was randomly split into a training set (90%) and an internal validation set (10%). An additional temporal model validation cohort was created from patients screened for DR during the calendar year 2021.

The outcomes of interest were PDR, refDR, and DME. Predictive models were developed and model performance evaluated by area under the curve (AUC) measures.

Baseline and time-varying clinical and demographic candidate predictors were collected from the KPNC EMR for each patient starting 12 months prior to baseline and continuing until the end of follow-up (when the participant either experienced an outcome of interest or was right-censored due to death, disenrollment, or the administrative end of the study).

Initial model development was based on 43 candidate nonocular clinical, demographic, and behavioral predictor variables, including lab results, diabetes medications, diagnoses (e.g., diabetic neuropathy), vital signs (pulse, respiratory rate, temperature, blood pressure), and resource utilization (Table 1). Some variables were observed at multiple points for each patient over the follow-up period preceding each subsequent DR screening visit. To minimize the complexity of the model and to simplify implementation, only the most recent data immediately preceding each DR screening encounter were used. When missing, continuous covariates were imputed using a median value with an additional indicator of missingness added as a separate covariate.

### Model Development

Extreme gradient boosting (XGBoost) and logistic regression models were trained, validated, and tested. XGBoost is an ML algorithm frequently favored by data scientists in predictive modeling studies, due to its efficiency, speed, and accuracy (19). Ninety percent of the cohort population, using data up to 31 December 2020, was used for training with fivefold cross-validation, and a 10% subset was used for internal validation. A second, temporal model validation and performance reporting step was done on a separate data set collected throughout the 2021 calendar year. Predictive models were developed for each of the three outcomes defined as onset of PDR, refDR, and DME.

To increase the usability, practicality, and transportability of our model in other health care settings, we developed a simplified version to remove covariates that may not be readily available to most practitioners, might introduce algorithmic bias, have high rates of missing values, or did not show good predictive performance in our ML models. The resulting LR models were trimmed to nine key covariates (Table 2).

### Statistical Analysis

For each model, we visually compared the receiver operating characteristic (ROC) curves for risk stratification of each modeling approach and valued their corresponding AUC with 95% CIs (20,21).

The resulting models were assessed for algorithmic bias, by evaluating the AUC performance of each model among protected subgroups defined by race and sex. We measured discrimination, the ability of a model to accurately distinguish between subjects who do versus do not develop the outcome, based on the AUC, with  $>0.7$  considered good discrimination. Calibration (the extent to which the predicted risks overestimate or underestimate the observed risks) was visually assessed using calibration plots. Given that the goal was not to quantify the numeric probability of an outcome for a given patient but rather to risk-stratify the population, model discrimination was prioritized over calibration. Statistical analysis was done in the R programming language (22). ML, training, and evaluation were performed with the h2o R package (23).

The study was reviewed and approved by the KPNC Institutional Review Board

**Table 1—Baseline characteristics of the cohorts**

|                                        | Full sample (N = 276,794)          |                                            |                                                  |
|----------------------------------------|------------------------------------|--------------------------------------------|--------------------------------------------------|
|                                        | Derivation cohort<br>(N = 249,148) | Internal validation cohort<br>(N = 27,646) | 2021 temporal validation<br>cohort (N = 169,678) |
| Sex (%)                                |                                    |                                            |                                                  |
| Female                                 | 119,917 (48.1)                     | 13,323 (48.2)                              | 81,233 (47.9)                                    |
| Male                                   | 129,217 (51.9)                     | 14,323 (51.8)                              | 88,439 (52.1)                                    |
| Other/unknown                          | 14 (0.0)                           | 0 (0.0)                                    | 6 (0.0)                                          |
| Race (%)                               |                                    |                                            |                                                  |
| White                                  | 101,891 (40.9)                     | 11,205 (40.5)                              | 113,096 (40.9)                                   |
| Asian                                  | 54,675 (21.9)                      | 6,185 (22.4)                               | 60,860 (22.0)                                    |
| Black                                  | 22,078 (8.9)                       | 2,459 (8.9)                                | 24,537 (8.9)                                     |
| Hispanic                               | 52,747 (21.2)                      | 5,769 (20.9)                               | 58,516 (21.1)                                    |
| Other                                  | 17,757 (7.1)                       | 2,028 (7.3)                                | 19,785 (7.1)                                     |
| Age in years at baseline               |                                    |                                            |                                                  |
| Mean (SD)                              | 60.0 (13.1)                        | 59.8 (13.2)                                | 64.1 (13.0)                                      |
| Median [min, max]                      | 61.0 [18.0, 102]                   | 61.0 [18.0, 100]                           | 65.0 [18.0, 103]                                 |
| Age in years at diabetes diagnosis     |                                    |                                            |                                                  |
| Mean (SD)                              | 55.9 (13.1)                        | 55.7 (13.2)                                | 54.0 (12.7)                                      |
| Median [min, max]                      | 56.0 [0, 103]                      | 56.0 [1.00, 99.0]                          | 55.0 [0, 99.0]                                   |
| Diabetes type (%)                      |                                    |                                            |                                                  |
| 1                                      | 4,858 (1.9)                        | 569 (2.1)                                  | 5,442 (3.2)                                      |
| 2                                      | 221,742 (89.0)                     | 24,512 (88.7)                              | 162,537 (95.8)                                   |
| Unknown                                | 22,548 (9.1)                       | 2,565 (9.3)                                | 1,699 (1.0)                                      |
| BMI, kg/m <sup>2</sup>                 |                                    |                                            |                                                  |
| Mean (SD)                              | 31.8 (6.73)                        | 31.8 (6.72)                                | 31.2 (6.93)                                      |
| Median [min, max]                      | 30.7 [10.6, 197]                   | 30.7 [14.5, 91.7]                          | 30.0 [10.4, 162]                                 |
| Systolic blood pressure, mmHg          |                                    |                                            |                                                  |
| Mean (SD)                              | 127 (14.2)                         | 127 (14.2)                                 | 129 (16.2)                                       |
| Median [min, max]                      | 126 [55.0, 258]                    | 126 [69.0, 227]                            | 130 [50.0, 248]                                  |
| Diastolic blood pressure, mmHg         |                                    |                                            |                                                  |
| Mean (SD)                              | 72.8 (9.82)                        | 72.8 (9.79)                                | 69.8 (11.8)                                      |
| Median [min, max]                      | 73.0 [0, 173]                      | 73.0 [0, 132]                              | 70.0 [20.0, 148]                                 |
| Hemoglobin A <sub>1c</sub> , %         |                                    |                                            |                                                  |
| Mean (SD)                              | 7.31 (1.63)                        | 7.35 (1.65)                                | 7.47 (1.57)                                      |
| Median [min, max]                      | 6.80 [3.50, 22.4]                  | 6.80 [4.30, 19.7]                          | 7.10 [4.20, 20.0]                                |
| Urinary albumin creatinine ratio, mg/g |                                    |                                            |                                                  |
| Mean (SD)                              | 19.6 (42.7)                        | 19.7 (43.5)                                | 45.4 (80.5)                                      |
| Median [min, max]                      | 8.70 [0.200, 1,050]                | 8.70 [0.200, 1,010]                        | 13.2 [0.700, 1,430]                              |
| Triglycerides, mg/dL                   |                                    |                                            |                                                  |
| Mean (SD)                              | 173 (172)                          | 173 (159)                                  | 172 (132)                                        |
| Median [min, max]                      | 145 [22.0, 13,400]                 | 145 [24.0, 5,400]                          | 145 [17.0, 9,540]                                |
| LDL, mg/dL                             |                                    |                                            |                                                  |
| Mean (SD)                              | 93.8 (31.8)                        | 93.9 (31.4)                                | 82.2 (33.2)                                      |
| Median [min, max]                      | 89.0 [1.00, 1,360]                 | 89.0 [15.0, 338]                           | 77.0 [7.00, 968]                                 |
| HDL, mg/dL                             |                                    |                                            |                                                  |
| Mean (SD)                              | 45.8 (10.9)                        | 45.8 (10.8)                                | 47.4 (11.7)                                      |
| Median [min, max]                      | 44.0 [4.00, 177]                   | 44.0 [5.00, 128]                           | 46.0 [4.00, 217]                                 |
| Creatinine, mg/g                       |                                    |                                            |                                                  |
| Mean (SD)                              | 0.913 (0.349)                      | 0.913 (0.356)                              | 1.00 (0.703)                                     |
| Median [min, max]                      | 0.870 [0.0700, 15.3]               | 0.870 [0.200, 14.0]                        | 0.890 [0.200, 20.8]                              |
| Smoking status (%)                     |                                    |                                            |                                                  |
| Current smoker                         | 14,141 (5.7)                       | 1,551 (5.6)                                | 7,464 (4.4)                                      |
| Former smoker                          | 52,285 (21.0)                      | 5,781 (20.9)                               | 49,719 (29.3)                                    |
| Passive smoker                         | 1,427 (0.6)                        | 169 (0.6)                                  | 654 (0.4)                                        |
| Never smoker                           | 115,487 (46.4)                     | 12,933 (46.8)                              | 102,121 (60.2)                                   |
| Unknown                                | 65,808 (26.4)                      | 7,212 (26.1)                               | 9,720 (5.7)                                      |

Continued on p. 1071

Table 1—Continued

|                                           | Full sample (N = 276,794)       |                                         | 2021 temporal validation cohort (N = 169,678) |
|-------------------------------------------|---------------------------------|-----------------------------------------|-----------------------------------------------|
|                                           | Derivation cohort (N = 249,148) | Internal validation cohort (N = 27,646) |                                               |
| UACR stage (%)                            |                                 |                                         |                                               |
| Macroalbuminuria                          | 455 (0.2)                       | 60 (0.2)                                | 1,386 (0.8)                                   |
| Microalbuminuria                          | 26,689 (10.7)                   | 2,964 (10.7)                            | 42,225 (24.9)                                 |
| Missing                                   | 105,139 (42.2)                  | 11,679 (42.2)                           | 26,489 (15.6)                                 |
| Normal                                    | 116,865 (46.9)                  | 12,943 (46.8)                           | 99,578 (58.7)                                 |
| Chronic kidney disease stage number (%)   |                                 |                                         |                                               |
| 1                                         | 97,620 (39.2)                   | 10,778 (39.0)                           | 64,265 (37.9)                                 |
| 2                                         | 101,162 (40.6)                  | 11,328 (41.0)                           | 72,604 (42.8)                                 |
| 3                                         | 26,500 (10.6)                   | 2,920 (10.6)                            | 25,614 (15.1)                                 |
| 4                                         | 1,183 (0.5)                     | 134 (0.5)                               | 2,232 (1.3)                                   |
| 5                                         | 336 (0.1)                       | 40 (0.1)                                | 1,524 (0.9)                                   |
| Missing                                   | 22,347 (9.0)                    | 2,446 (8.8)                             | 3,439 (2.0)                                   |
| Diabetic neuropathy                       | 23,261 (9.3)                    | 2,488 (9.0)                             | 40,858 (24.1)                                 |
| Diabetes medications (%) <sup>a</sup>     |                                 |                                         |                                               |
| Metformin                                 | 148,500 (59.6)                  | 16,482 (59.6)                           | 114,654 (67.6)                                |
| Sulfonylurea                              | 113,607 (45.6)                  | 12,635 (45.7)                           | 70,115 (41.3)                                 |
| Insulin                                   | 65,848 (26.4)                   | 7,388 (26.7)                            | 50,182 (29.6)                                 |
| Thiazolidinediones                        | 23,512 (9.4)                    | 2,612 (9.4)                             | 7,564 (4.5)                                   |
| Dipeptidyl peptidase 4 inhibitors         | 4,355 (1.7)                     | 472 (1.7)                               | 3,048 (1.8)                                   |
| Glucagon-like peptide-1                   | 1,563 (0.6)                     | 191 (0.7)                               | 2,626 (1.5)                                   |
| Sodium-glucose cotransporter-2 inhibitors | 1,481 (0.6)                     | 156 (0.6)                               | 9,258 (5.5)                                   |
| Other                                     | 1,263 (0.5)                     | 143 (0.5)                               | 278 (0.2)                                     |
| None                                      | 77,475 (31.1)                   | 8,559 (31.0)                            | 32,267 (19.0)                                 |
| Positive outcomes (%)                     |                                 |                                         |                                               |
| PDR                                       | 2,434 (1.0)                     | 279 (1.0)                               | 1,350 (0.8) <sup>b</sup>                      |
| refDR                                     | 17,108 (6.9)                    | 1,970 (7.1)                             | 6,284 (4.2) <sup>c</sup>                      |
| DME                                       | 7,404 (3.0)                     | 825 (3.0)                               | 1,828 (1.2) <sup>d</sup>                      |

For overall cohort, baseline is the date of visit when no DR is confirmed by fundus photograph; for 2021 cohort, baseline is 1 January 2021. All clinical inputs are last measurement prior to baseline. min, minimum value; max, maximum value. <sup>a</sup>Medication dispensed during 6 months prior to baseline; <sup>b</sup>n for validation cohort = 161,819; <sup>c</sup>n for validation cohort = 145,056; <sup>d</sup>n for validation cohort = 158,094.

(Oakland, CA) and adhered to the tenets of the Declaration of Helsinki.

## RESULTS

The study population included 276,794 patients; 51.9% were male, and 42.2% were White (Table 1). The mean age was 60.0 ± 13.1 years. 89% had type 2 diabetes, and 26.4% used insulin. Approximately 1.0% of the population eventually became positive for PDR, 6.9% for refDR, and 3.0% for DME.

### Model Performance

Both the full 43 covariate ML and 9 covariate (trimmed) LR models performed best when predicting the PDR outcome (ML AUC 0.86; 95% CI 0.86–0.87) versus (LR AUC 0.82; 95% CI 0.80–0.83). ROC curves and AUCs from the 2021 temporal validation and internal validation data sets are shown in Fig. 1. The ML model including 43 variables slightly outperformed the trimmed LR model using

only 9 variables for each of the three outcome measures. Both approaches were able to accurately predict refDR and DME outcomes.

At 80% screening capacity, in the 2021 temporal validation data set, the LR model for PDR outcome had 98.3% sensitivity, 20.2% specificity, 0.01 positive predictive value (PPV), 0.999 negative predictive value; 128,128 type 1 errors and 23 type 2 errors (Supplementary Table 1).

The nine variable LR models were examined for algorithmic bias for selected patient subgroups determined by race and sex, and comparable AUCs were found for all groups, suggesting algorithmic fairness.

### Odds Ratios

Odds ratios (OR) were calculated to estimate the clinical contribution of the nine variables in the LR model (Table 2). Insulin use made the largest contribution for

PDR (OR 3.768, 95% CI 3.439–4.129), refDR (OR 2.996, 95% CI 2.897–3.098), and DME (OR 2.727, 95% CI 2.596–2.865). Other variables also showed significant effects, notably, evidence of diabetic neuropathy and elevated creatinine (which are generally indicative of end-organ damage). Increased BMI and age were slightly protective for all three outcomes.

### Sensitivity

The sensitivity of our LR model was calculated and performance plotted over a range of screening capacities of 10% to 90% (Fig. 2). As an example, in the 2021 temporal cohort, there were 1,350 incident PDR cases identified. If random screening was performed at 50% screening capacity, 675 cases of PDR would have been identified, and another 675 cases of PDR would not have been screened. If the cohort was first risk stratified by the LR model and then screened, only 148 cases

**Table 2—OR from multivariate LR models for PDR, refDR, and DME**

| Outcome                              | $\beta$ | OR    | 95% CI      |
|--------------------------------------|---------|-------|-------------|
| <b>PDR</b>                           |         |       |             |
| Covariate                            |         |       |             |
| Insulin (0/1)                        | 1.333   | 3.768 | 3.439–4.129 |
| HgbA <sub>1c</sub> (%)               | 0.150   | 1.162 | 1.140–1.185 |
| Age at visit (years)                 | −0.030  | 0.966 | 0.964–0.969 |
| Systolic BP (mmHg)                   | 0.010   | 1.010 | 1.008–1.013 |
| BMI (kg/m <sup>2</sup> )             | −0.020  | 0.975 | 0.969–0.980 |
| Diabetic peripheral neuropathy (0/1) | 0.743   | 2.104 | 1.936–2.286 |
| LDL (mg/dL)                          | 0.002   | 1.002 | 1.002–1.004 |
| UACR (mg/g)                          | 0.003   | 1.003 | 1.003–1.004 |
| Serum creatinine (mg/dL)             | 0.205   | 1.238 | 1.195–1.263 |
| <b>refDR</b>                         |         |       |             |
| Covariate                            |         |       |             |
| Insulin (0/1)                        | 1.097   | 2.996 | 2.897–3.098 |
| HgbA <sub>1c</sub> (%)               | 0.251   | 1.285 | 1.275–1.296 |
| Age at visit (years)                 | −0.222  | 0.978 | 0.977–0.979 |
| Systolic BP (mmHg)                   | 0.011   | 1.011 | 1.010–1.012 |
| BMI (kg/m <sup>2</sup> )             | −0.023  | 0.976 | 0.975–0.979 |
| Diabetic peripheral neuropathy (0/1) | 0.349   | 1.418 | 1.369–1.468 |
| LDL (mg/dL)                          | 0.002   | 1.002 | 1.002–1.003 |
| UACR (mg/g)                          | 0.002   | 1.002 | 1.002–1.003 |
| Serum creatinine (mg/dL)             | 0.139   | 1.149 | 1.127–1.172 |
| <b>DME</b>                           |         |       |             |
| Covariate                            |         |       |             |
| Insulin (0/1)                        | 1.003   | 2.727 | 2.596–2.865 |
| HgbA <sub>1c</sub> (%)               | 0.123   | 1.130 | 1.116–1.145 |
| Age at visit (years)                 | 0.005   | 1.005 | 1.003–1.006 |
| Systolic BP (mmHg)                   | 0.009   | 1.009 | 1.007–1.010 |
| BMI (kg/m <sup>2</sup> )             | −0.013  | 0.986 | 0.983–0.990 |
| Diabetic peripheral neuropathy (0/1) | 0.376   | 1.457 | 1.389–1.529 |
| LDL (mg/dL)                          | 0.004   | 1.004 | 1.004–1.005 |
| UACR (mg/g)                          | 0.002   | 1.002 | 1.002–1.003 |
| Creatinine (mg/dL)                   | 0.120   | 1.127 | 1.102–1.154 |

Nonstandardized ORs from multivariate LR models of all nine covariates were calculated for each outcome. The 95% CIs are included. All *P* values were <0.001. BP, blood pressure; HgbA<sub>1c</sub>, glycosylated hemoglobin.

of PDR would have been missed. This means a prioritized, model-based approach reduces the missed cases by nearly 80% compared with random screening at a 50% capacity.

### Calibration

Model calibration was assessed visually by plotting the observed event rate against the predicted event rate based on deciles of predicted risk. For all three outcomes (Supplementary Fig. 1), our model slightly overestimates the number of events, thus limiting the chance for false negatives. As previously discussed, the objective of our models was discrimination and not calculation of continuous probability risk.

### Type 1 Diabetes

Diabetes is classified as type 1 or 2, although the ocular management of patients

with retinopathy is similar in both groups. Even though type 1 patients only represent less than 5% of the total population, we were able to show there was no algorithmic bias for subtype of diabetes (Supplementary Fig. 2).

### CONCLUSIONS

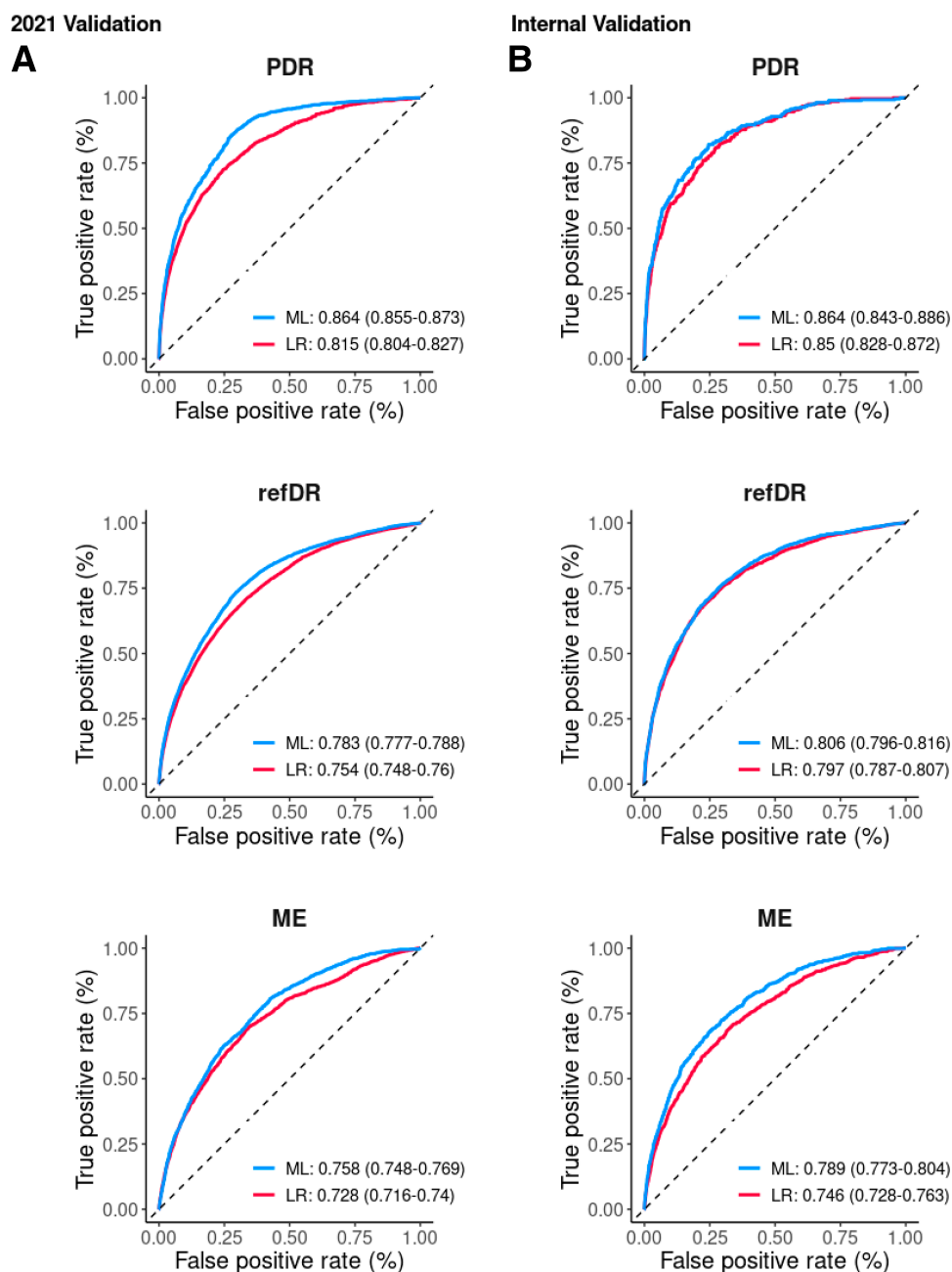
This study compared ML and LR risk stratification models for onset of PDR, refDR, and DME. Both approaches are able to accurately predict the onset of advanced stages of DR. Surprisingly, a trimmed LR model using only nine variables proved to be nearly as effective as the 43-variable ML model. The nine-variable model provided 80% more sensitivity for PDR at 50% screening capacity compared with random screening. Sensitivity for the other outcomes of DME and refDR were similarly high.

A first step in effective population management is risk stratification. In many health systems in the U.S., DR screening is driven purely by compliance with Healthcare Effectiveness Data and Information Set screening guidelines. Patients are screened using a “one size fits all” algorithm. Often, screening efforts are incomplete due to constraints in capacity or patient compliance. The COVID-19 pandemic has adversely impacted screening access and further contributed to challenges in DR screening. Unfortunately, those who are overdue for DR screening are at highest risk for complications. Especially for the outcome of PDR, failure to identify disease in a timely manner can lead to severe vision loss.

One approach to capture patients at higher risk for adverse outcomes in the face of limited screening resources is to rank order patients due for screening, based on estimated clinical risk. Our results suggest that, especially at lower screening capacities, the risk of missing a PDR diagnosis would be significantly reduced using our simplified LR model to identify the highest-risk population for screening prioritization. While more complicated ML models can be more accurate in predicting adverse outcomes, there is a cost of the added complexity they provide. These models are often “black box,” without the transparency to understand the importance of individual variables, and also are typically more difficult to implement.

This study is the largest to date to predict DR strictly based on nonocular, systemic risk factors. Recent studies have been smaller (24,25) or have used both clinical records and fundus photography strategies (26) to train and validate their models. Similar to these studies, we found the most important covariate to be insulin dependence. This is not surprising since, fundamentally, diabetes is a disease of blood glucose metabolism deregulation.

Other strengths of this study included up to 12 years of follow-up, enabling capture of more clinical events and rare outcomes, like PDR, over time. Despite the very low prevalence of some outcomes, the models were able to perform at very high sensitivities at multiple screening capacities. Using these models to identify patients at highest risk and provide timely intervention to prevent



**Figure 1**—Performance of prediction models on validation cohort data. ROC plots of XGBoost (ML) and LR for three outcome measures: PDR, refDR, and DME in a temporal validation cohort (A) and internal validation cohort (B). Additional results report AUC and 95% CI for each outcome.

further disease progression should lead to improved quality of care.

Significant limitations of this study include its retrospective nature. Our study strictly considered systemic, nonophthalmological clinical risk factors for retinopathy onset. Duration of diabetes is recognized as a significant risk factor for DR onset and progression (27). That said, date of detection of diabetes may not be readily available in the EMR in most health care settings. Since our primary objective was to create a model that will be readily transportable to other

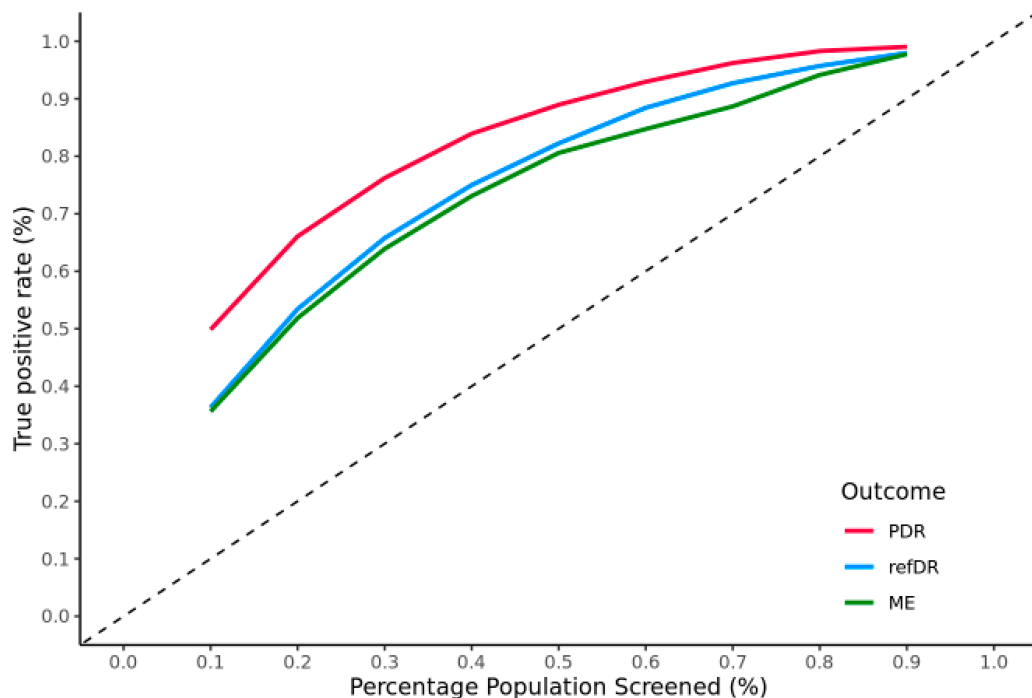
health care systems, we purposely excluded variables not readily available and thus did not include duration of diabetes as a covariate in our analysis.

Local mediators of neovascularization such as posterior vitreous separation status, vascular endothelial growth factor (VEGF), and intraocular cytokine levels can influence retinopathy outcomes. These were not included in our analysis.

The urine albumin to creatinine ratio (UACR) is predictive in our model. UACR is a ratio of measured albumin and creatinine in urine and serves as a marker

of nephropathy (28). The American Diabetes Association has recommended yearly screening for patients with type 1 diabetes (disease over 5 years duration) and those with type 2 diabetes, but this is rarely achieved in clinical practice (29). In our cohort, approximately 40% of patients had UACR measurements. While the low rate of UACR measurement in our model could influence performance, our model still functions quite well.

Our models perform with very high sensitivity and high negative predictive value. However, since our measured



**Figure 2**—Sensitivity of LR model compared with random screening. At all screening capacities, the model-based approach outperforms random screening (diagonal line) for all three outcome measures.

outcomes were rare (prevalence <5%), specificity and PPV were low. The purpose of our models is to rank order a population due for screening but not influence clinical decision-making or measure risk on an individual basis. Thus, lower PPV have no substantive clinical consequences.

Patients overdue for screening are at higher risk, since their retinopathy status is unknown. Post model implementation validation studies can be done to compare model performance in patients that were screened on time versus those that were screened when overdue.

Other limitations included reliance on data from the KPNC EMR and validation in the same single U.S. health care delivery system that the models were derived from. Though our models performed well, external validation should be done in other health care systems, as their performance may be less robust when using data sets from other health systems with less comprehensive or reliable input capture.

In summary, a simple and practical model was developed that can rank order the risk of retinopathy in a population with diabetes that has been previously screened and is now overdue, facilitating prioritizing screening for retinopathy

complications among those at highest risk. Clinical researchers may find this tool helpful in identifying patients at high risk for diabetic eye disease, for either purposeful inclusion or exclusion in clinical trials of novel therapies and diagnostic tests. Quality improvement and implementation studies are needed to evaluate whether and how this retinopathy risk stratification tool may influence provider behavior, screening, and rates of diabetic eye disease. Health care delivery systems can use this model to allocate resources to improve outreach and compliance with the highest-risk populations. Future research can identify optimal screening frequencies for lower-risk populations, yielding higher efficiency and safety for all patients with diabetes.

**Funding.** This project was supported by The Permanente Medical Group Delivery Science Grants Program (RNG 211024). A.J.K. was also supported by National Institute of Diabetes and Digestive and Kidney Diseases Centers for Diabetes Translation Research (P30 DK092924).

Funders were not involved with the design or conduct of the study, collection, management, analysis, or interpretation of the data, preparation, review, or approval of the manuscript, or decision to submit the manuscript for publication.

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

**Author Contributions.** D.T., A.J.K., H.H.M., O.S., and R.B.M. made substantial contributions to the conception and design. K.K.T. and D.S. made substantial contributions to the analysis and interpretation of data. D.T. and O.S. drafted the manuscript. A.J.K., H.H.M., and R.B.M. provided critical revision of the manuscript for important intellectual content. N.P., K.K.T., D.S., and O.S. performed statistical analysis. D.T. and O.S. obtained funding. D.T. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## References

- Unnikrishnan R, Pradeepa R, Joshi SR, Mohan V. Type 2 diabetes: demystifying the global epidemic. *Diabetes* 2017;66:1432–1442
- International Diabetes Foundation. *IDF Diabetes Atlas*. 9th ed. International Diabetes Foundation, 2019
- American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. *Diabetes Care* 2013;36:1033–1046
- Wilkinson CP, Ferris FL 3rd, Klein RE, et al.; Global Diabetic Retinopathy Project Group. Proposed international clinical diabetic retinopathy and diabetic macular edema disease severity scales. *Ophthalmology* 2003;110:1677–1682
- Division of Diabetes Translation, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention. Vision Health Initiative, 2020. Accessed 11 March 2023. Available from <https://www.cdc.gov/visionhealth/risk/tips.htm>



6. Yau JW, Rogers SL, Kawasaki R, et al.; Meta-Analysis for Eye Disease (META-EYE) Study Group. Global prevalence and major risk factors of diabetic retinopathy. *Diabetes Care* 2012;35:556–564
7. Lee R, Wong TY, Sabanayagam C. Epidemiology of diabetic retinopathy, diabetic macular edema and related vision loss. *Eye Vis (Lond)* 2015;2:17
8. Williams GA, Scott IU, Haller JA, Maguire AM, Marcus D, McDonald HR. Single-field fundus photography for diabetic retinopathy screening: a report by the American Academy of Ophthalmology. *Ophthalmology* 2004;111:1055–1062
9. Lin DY, Blumenkranz MS, Brothers RJ, Grosvenor DM. The sensitivity and specificity of single-field nonmydriatic monochromatic digital fundus photography with remote image interpretation for diabetic retinopathy screening: a comparison with ophthalmoscopy and standardized mydriatic color photography. *Am J Ophthalmol* 2002;134:204–213
10. American Academy of Ophthalmology. *Diabetic Retinopathy Preferred Practice Pattern*. American Academy of Ophthalmology, 2019
11. Solomon SD, Chew E, Duh EJ, et al. Diabetic retinopathy: a position statement by the American Diabetes Association. *Diabetes Care* 2017;40:412–418
12. Chin EK, Ventura BV, See KY, Seibles J, Park SS. Nonmydriatic fundus photography for tele-ophthalmology diabetic retinopathy screening in rural and urban clinics. *Telemed J E Health* 2014;20:102–108
13. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med* 2018;1:39
14. Ipp E, Liljenquist D, Bode B, et al.; EyeArt Study Group. Pivotal evaluation of an artificial intelligence system for autonomous detection of referable and vision-threatening diabetic retinopathy. *JAMA Netw Open* 2021;4:e2134254
15. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin epidemiologic study of diabetic retinopathy. II. Prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years. *Arch Ophthalmol* 1984;102:520–526
16. Klein R, Klein BE, Moss SE, Davis MD, DeMets DL. The Wisconsin epidemiologic study of diabetic retinopathy. III. Prevalence and risk of diabetic retinopathy when age at diagnosis is 30 or more years. *Arch Ophthalmol* 1984;102:527–532
17. Klein R, Knudtson MD, Lee KE, Gangnon R, Klein BE. The Wisconsin Epidemiologic Study of Diabetic Retinopathy XXII: the twenty-five-year progression of retinopathy in persons with type 1 diabetes. *Ophthalmology* 2008;115:1859–1868
18. Moffet HH, Adler N, Schillinger D, et al. Cohort profile: The Diabetes Study of Northern California (DISTANCE) objectives and design of a survey follow-up of social health disparities in a managed care population. *Int J Epidemiol* 2009;38:38–47
19. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, 13–17 August 2016*. New York, Association for Computing Machinery, pp. 785–794
20. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36
21. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. New York, Oxford University Press, 2003
22. R Core Team. R: A Language and Environment for Statistical Computing. Accessed 11 March 2023. Available from <http://www.r-project.org/index.html>
23. H2O.ai. R Interface for H2O, R package version 3.36.0.4, 2022. Accessed 11 March 2023. Available from <https://github.com/h2oai/h2o-3>
24. Oh E, Yoo TK, Park EC. Diabetic retinopathy risk prediction for fundus examination using sparse learning: a cross-sectional study. *BMC Med Inform Decis Mak* 2013;13:106
25. Ogunyemi OI, Gandhi M, Lee M, et al. Detecting diabetic retinopathy through machine learning on electronic health record data from an urban, safety net healthcare system. *JAMIA Open* 2021;4:b066
26. Bora A, Balasubramanian S, Babenko B, et al. Predicting the risk of developing diabetic retinopathy using deep learning. *Lancet Digit Health* 2021;3:e10–e19
27. Fong DS, Aiello L, Gardner TW, et al.; American Diabetes Association. Diabetic retinopathy. *Diabetes Care* 2003;26:226–229
28. Gross JL, de Azevedo MJ, Silveiro SP, Canani LH, Caramori ML, Zelmanovitz T. Diabetic nephropathy: diagnosis, prevention, and treatment. *Diabetes Care* 2005;28:164–176
29. Stempniewicz N, Vassalotti JA, Cuddeback JK, et al. Chronic kidney disease testing among primary care patients with type 2 diabetes across 24 U.S. health care organizations. *Diabetes Care* 2021;44:2000–2009