AMERICAN ACADEMY OF OPHTHALMOLOGY®
In partnership with the American Glaucoma Society

# Validating and Updating the OHTS-EGPS Model Predicting 5-year Glaucoma Risk among Ocular Hypertension Patients Using Electronic Records

*David M. Wright, PhD,[1,*] Augusto Azuara-Blanco, PhD,[1,2,*] Chris Cardwell, PhD,[1] Giovanni Montesano, PhD,[3] David P. Crabb, PhD,[3] Gus Gazzard, MD,[4] Anthony J. King, MD,[5] Rodolfo Hernández, PhD,[6] James E. Morgan, DPhil,[7] Bethany Higgins, PhD,[1,3] Yemisi Takwoingi, PhD,[8] on behalf of the GRIP study group*

***Objective:*** To validate and update the Ocular Hypertension Treatment Study-European Glaucoma Prevention Study (OHTS-EGPS) model predicting risk of conversion from ocular hypertension (OHT) to glaucoma using electronic medical records (EMR).

***Design:*** Evaluation and update of a risk prediction algorithm using EMRs and linked visual field (VF) tests.

***Participants:*** Newly diagnosed OHT patients attending hospital glaucoma services in England. Inclusion criteria are as follows: intraocular pressure (IOP) 22 to 32 mmHg (either eye); normal baseline VF test, defined as Glaucoma Hemifield Test (GHT) "within normal range" in a reliable VF test; at least 2 VF tests in total; no significant ocular comorbidities.

***Methods:*** Risk factors are as follows: age, ethnicity, sex, IOP, vertical cup-to-disc ratio, central corneal thickness, VF pattern standard deviation, family history of glaucoma, systemic hypertension, diabetes mellitus, and glaucoma treatment. Glaucoma conversion was defined as 2 consecutive and reliable VF tests with GHT "outside normal limits" and/or need for glaucoma surgery. For validation, the OHTS-EGPS model was applied to predict a patient's risk of developing glaucoma in 5 years. In the updating stage, the OHTS model was refitted by re-estimating the baseline hazard and regression coefficients. The updated model was cross-validated and several variants were explored.

***Main Outcome Measures:*** Measures of discriminative ability (c-index) and calibration (calibration slope) were calculated and pooled across hospitals using random effects meta-analysis.

***Results:*** From a total of 138 461 patients from 10 hospital glaucoma services in England, 9030 patients with OHT fitted the inclusion criteria. A total of 1530 (16.9%) patients converted to glaucoma during this follow-up period. The OHTS-EGPS model provided a pooled c-index of 0.61 (95% confidence interval: 0.60−0.63), ranging from 0.55 to 0.67 between hospitals. The pooled calibration slope was 0.45 (0.38−0.51), ranging from 0.25 to 0.64 among hospitals. The overall refitted model performed better than the OHTS-EGPS model, with a pooled c-index of 0.67 (0.65−0.69), ranging from 0.65 to 0.75 between hospitals.

***Conclusions:*** We performed an external validation of the OHTS-EGPS model in a large English population. Refitting the model achieved modest improvements in performance. Given the poor performance of the OHTS-EGPS model in our population, one should use caution in its application to populations that differ from those in the OHTS and EGPS.

***Financial Disclosure(s):*** Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Glaucoma 2025;8:143-151 © 2024 by the American Academy of Ophthalmology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).*

*Supplemental material available at www.ophthalmologyglaucoma.org.*

Glaucoma remains a major cause of vision loss worldwide,[1] with an estimated 4.1 million cases with moderate or severe vision impairment in 2020,[2] and glaucoma prevalence expected to reach 112 million by 2040.[3] Ocular hypertension (OHT), defined as intraocular pressure > 21 mmHg and with a normal optic disc and visual field (VF), is a major risk factor for glaucoma.[4−7]

Monitoring the growing number of OHT patients threatens to overwhelm glaucoma services. In the UK, around 1.3 million people aged over 40 have OHT, with a



https://doi.org/10.1016/j.ogla.2024.10.009
ISSN 2589-4196/24

16% increase in this population expected by 2035.[8] Standard practice is to monitor OHT in hospitals; guidelines recommend regular monitoring visits and treatment according to risk. However, only a small proportion of OHT patients progress to glaucoma each year. The development of a tool that can usefully predict the risk of developing glaucoma in patients with OHT has been identified as a key research priority.[9] This would enable clinicians to prioritize resources and recommend more frequent monitoring and treatment to those at highest risk. A risk calculator based on the OHTS-EGPS studies is available,[10] but it is not recommended in clinical guidelines.

In this study, we aimed to validate and update the OHTS-EGPS model on the risk of conversion to glaucoma using electronic medical records (EMRs) of a large cohort from 10 hospitals in England. Specifically, we sought to optimize tools for the prediction of the 5-year risk of glaucoma onset in a diverse population with OHT.

## Methods

### Population

We included adults with newly diagnosed OHT in one or both eyes, as recorded in the EMR. Ocular hypertension was defined as intraocular pressure (IOP) $\geq$ 22 mmHg and $\leq$ 32 mmHg measured using Goldmann applanation tonometry, no clinical signs of primary open angle glaucoma (POAG, i.e., normal optic nerve examination and normal VF test), and no associated abnormalities on clinical examination (e.g., pigment dispersion or pseudoexfoliation). “Normal” VF was defined as a reliable standard automated perimetry with Humphrey visual fields with a Glaucoma Hemifield Test (GHT) “within normal limits.”

We excluded eyes with clinically significant ocular comorbidity, such as maculopathies, and patients with glaucoma (any type) in one eye at baseline. To match the original OHTS study,[4] those with IOP > 32 mmHg in either eye at baseline were excluded as “glaucoma suspects.” We excluded those who did not have any VF testing and those without reliable VF testing. An unreliable VF was defined as a high frequency of false positives, > 15%.[11] The unit of analysis was the person. Some patients contributed only one eye to the analysis if the other eye was excluded due to an ocular comorbidity.

### Data Extraction

EMR data were extracted for 10 hospitals in England (listed in acknowledgments) that used the Medisoft platform (Medisoft). All hospitals were state-funded, part of the UK National Health Service. Hospitals were selected to provide sufficient statistical power, given the population sizes and number of glaucoma conversions expected based on previous EMR analyses.[12] This study adheres to the tenets of the Declaration of Helsinki. Ethical approval for use of these data was obtained (REC reference 21/EE/0109) and permissions received from each center. Because all records were anonymized prior to extraction, patient consent was not required. The study protocol has been published.[13] The Medisoft platform is based around a relational database containing tables for each type of record (e.g., patient demographics, clinical encounters, IOP measurements). The database is populated through a graphical interface with text boxes and drop-down menus, which have defined data fields that must be correctly completed before the record can be saved. (Fig S1, available at www.ophthalmologyglaucoma.org/). This structured data collection approach reduces the probability of data entry errors. Visual fields were captured at each site using the Humphrey Field Analyzer and automatically imported into the Medisoft platform. Visual field measurements were included in the main data extraction and comprised both global measures (e.g., Glaucoma Hemifield Test, false positive rate) and pointwise sensitivity measurements. For this study, records for all patients with a diagnosis of ocular hypertension or glaucoma at any timepoint in the hospital databases were batch extracted with the assistance of Medisoft. Before extraction, all personal identifiers were removed, and visit dates and dates of birth were perturbed (+-180 days) to preserve patient confidentiality.

### Cohort Preparation

Data were extracted for patients (n = 138 461) who attended any of the 10 hospitals between November 1995 and January 2022 (Fig 2). After exclusions, the analysis dataset comprised 9030 patients (13 891 eyes).

### Statistical Analysis

**Outcome.** The primary outcome was conversion to glaucoma within 5 years. We used VF tests to detect conversion, defined as 2 consecutive reliable abnormal VF examination results (i.e., GHT outside normal limits[12]). The date of conversion was the date of the first abnormal result. If an eye underwent surgery for glaucoma, even in the absence of VF conversion according to our definition, conversion was deemed to have occurred (at the earlier date). Patients were followed from the first normal VF test (after OHT diagnosis) until the date of glaucoma conversion or censored at 5 years after the first normal VF test, visual acuity dropping below 6/18, or diagnosis of an ocular co-pathology affecting VF (whichever was earliest). Eyes were not excluded or censored based on a diagnosis of diabetic retinopathy because severity stages were not consistently recorded and the dataset would have included many background cases that exerted little influence on VFs. Changes to vertical cup-to-disc ratio (VCDR) measurements were not used to determine conversion to glaucoma.

**Predictors.** Data extracted are listed in Table 1. Values outside the range of predictors in the OHTS-EGPS cohort were ignored (considered missing).[14] Patients included in the cohort were newly diagnosed (i.e., they were not under treatment). Some patients started treatment at time of diagnosis. The predictors included the treatment status at OHT diagnosis (received IOP-lowering medical treatment, yes or no). Details of data preparation are given in the Electronic Supplementary Materials, available at www.ophthalmologyglaucoma.org/).

**Validation of the OHTS-EGPS Model.** The original OHTS-EGPS risk prediction model (model A) was applied to all 10 hospital datasets to calculate the predicted risk of developing glaucoma in 5 years for each participant as previously described[14]:

OHTS-EGPS predicted risk $= 1 - 0.92^{\exp(PI)}$

$PI = 0.23\times(\text{age}^1 - 5.64) + 0.09\times(\text{IOP} - 24.13) + 0.71\times(\text{CCT}^2 + 14.33) + 0.12\times(\text{PSD}^3 - 9.76) + 0.18\times(\text{VCDR}^4 - 3.60))$

$^1$age in decades; $^2$per reduction of 40μm central corneal thickness (CCT); $^3$per increase of 0.2 pattern standard deviation (PSD); $^4$per increase of 0.1 VCDR.

Model discrimination was assessed using Harrell c-index and calibration using the calibration slope. The c-index measures how well predicted risk scores describe the observed sequence of events: the probability that a randomly selected pair of patients are ordered correctly. A c-index of 1.0 indicates that the risk score ordered all patients correctly; a score of 0.5 indicates ordering no

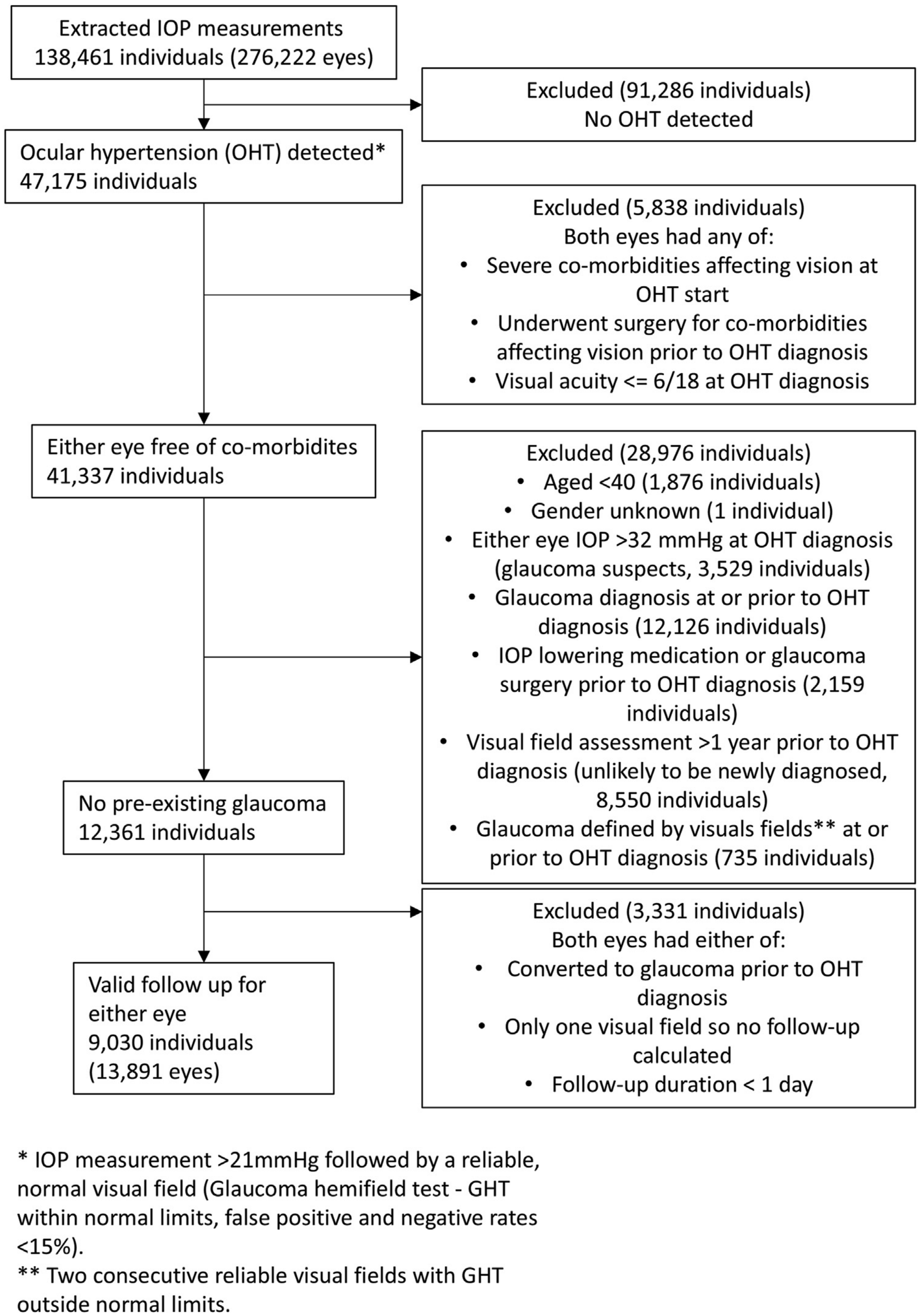


**Figure 1.** Flowchart describing construction of analysis cohort. IOP = intraocular pressure; OHT = ocular hypertension; GHT = Glaucoma Hemifield Test.

better than random (further details in the Electronic Supplementary Materials).

The calibration slope measures how closely the predicted risk matches the observed risk. Calibration plots of average observed risk against predicted 5-year risk were used to assess calibration at each hospital. Participants were grouped by predicted risk, and the average predicted risk for each group was compared with the corresponding Kaplan-Meier estimate of the observed risk within each hospital. Quintiles were used to ensure sufficient data support for stable estimates of risk in each group.

C-index and calibration were calculated for each hospital and pooled across hospitals using random effects meta-analysis. The calibration slope was pooled on the original scale, and the c-index was transformed to the logit scale before meta-analysis.[15]

There were moderate proportions of missing CCT, vertical cup-to-disc ratio (VCD), and PSD measurements (Table 1). To minimize the risk of bias due to missing data, values were imputed using multiple imputation by chained equations, a widely used technique that generates imputed datasets to account

Table 1. Characteristics of Participants with Glaucoma within 5 Years and Those with No Glaucoma within 5 Years (Not Restricted Based Upon 5 Years of Follow-up)

| | | Entire Cohort | |
|---|---|---|---|
| **Variable** | **Category** | *No Glaucoma* | *Glaucoma* |
| N | | 7500 | 1530 |
| Age, mean (SD) | | 61.5 (10.5) | 65.6 (10.4) |
| Age, yrs | 40–49 | 1249 (17%) | 132 (9%) |
| | 50–59 | 2070 (28%) | 321 (21%) |
| | 60–69 | 2488 (33%) | 505 (33%) |
| | 70–79 | 1421 (19%) | 468 (31%) |
| | ≥ 80 | 272 (4%) | 104 (7%) |
| Male | | 4084 (54%) | 840 (55%) |
| Hospital ID | 1 | 447 (6%) | 117 (8%) |
| | 2 | 366 (5%) | 64 (4%) |
| | 3 | 537 (7%) | 91 (6%) |
| | 4 | 337 (4%) | 111 (7%) |
| | 5 | 996 (13%) | 159 (10%) |
| | 6 | 2165 (29%) | 512 (33%) |
| | 7 | 1084 (14%) | 251 (16%) |
| | 8 | 758 (10%) | 118 (8%) |
| | 9 | 621 (8%) | 82 (5%) |
| | 10 | 189 (3%) | 25 (2%) |
| IOP, mean (SD) | | 25.0 (2.6) | 25.1 (2.7) |
| IOP, mmHg | < 22.5 | 1302 (17%) | 289 (19%) |
| | 22.5–25 | 2588 (35%) | 474 (31%) |
| | 25–27.5 | 2099 (28%) | 453 (30%) |
| | 27.5–30 | 988 (13%) | 193 (13%) |
| | ≥ 30 | 523 (7%) | 121 (8%) |
| CCT, mean (SD) | | 560.3 (35.6) | 553.0 (35.1) |
| CCT, µm | < 500 | 278 (4%) | 74 (5%) |
| | 500–549 | 2214 (30%) | 511 (33%) |
| | 550–599 | 2953 (39%) | 538 (35%) |
| | ≥ 600 | 869 (12%) | 112 (7%) |
| | Missing | 1186 (16%) | 295 (19%) |
| PSD, mean (SD) | | 1.6 (0.3) | 1.8 (0.4) |
| PSD, dB | < 1.5 | 2186 (29%) | 215 (14%) |
| | 1.5-2 | 2268 (30%) | 497 (32%) |
| | 2-2.5 | 485 (6%) | 178 (12%) |
| | ≥ 2.5 | 97 (1%) | 44 (3%) |
| | Missing | 2464 (33%) | 596 (39%) |
| VCDR, mean (SD) | | 0.5 (0.2) | 0.5 (0.2) |
| VCDR | < 0.2 | 169 (2%) | 23 (2%) |
| | 0.2–0.4 | 1286 (17%) | 179 (12%) |
| | 0.4–0.6 | 1873 (25%) | 313 (20%) |
| | 0.6–0.8 | 1248 (17%) | 301 (20%) |
| | ≥ 0.8 | 103 (1%) | 40 (3%) |
| | Missing | 2821 (38%) | 674 (44%) |
| Ethnicity | White | 4796 (64%) | 1076 (70%) |
| | Non-White | 487 (6%) | 118 (8%) |
| | Not stated | 2217 (30%) | 336 (22%) |
| FH glaucoma | | 2002 (27%) | 368 (24%) |
| Diabetes | | 973 (13%) | 265 (17%) |
| Hypertension | | 1059 (14%) | 173 (11%) |
| Treatment | | 2220 (30%) | 502 (33%) |

CCT = central corneal thickness; FH = family history; IOP = intraocular pressure; PSD = pattern standard deviation; SD = standard deviation; VCDR = vertical cup-to-disc ratio.

for the uncertainty associated with missing values.[16] Ten imputations were applied.[17] The imputation model was stratified by hospital and included all OHTS-EGPS predictors, the event indicator, and cumulative hazard. Estimates were pooled across imputations using Rubin rules. The analysis was repeated using complete cases as a sensitivity analysis.

**Updating the OHTS-EGPS Model.** An updated model (model B) was fitted using the OHTS-EGPS predictors but re-estimating the baseline hazard and the regression coefficients to improve both calibration and discrimination. An internal-external cross-validation was used, developing the model using data from 9 hospitals and validating in the remaining hospital. This was repeated 10 times with c-index and calibration slopes pooled by meta-analysis.

**Further Model Variants Were Explored.** Model C includes all additional predictors except ethnicity (due to missing data), model D uses the IOP of the worse eye (i.e., eye with the highest IOP at baseline) to investigate the impact of averaging IOP across eyes on model B, model E includes only patients with IOP ≤ 23 mmHg, model F includes only patients who had not received IOP treatments at baseline, and model G includes only patients who never received IOP treatments.

### Influence of Baseline Variables on Treatment

Treatment with IOP-lowering medication may influence the probability of conversion from OHT to glaucoma and hence the performance of risk prediction models. The decision to start treatment with IOP-lowering medication for those with OHT is largely dependent on a small number of clinical characteristics: age, family history of glaucoma, CCT, and IOP. To set the risk prediction models in context, we modeled the associations between these variables and probability of having received IOP-lowering medication at baseline using logistic regression.

## Results

### Validation of the OHTS-EGPS Model

A total of 1530 of 9030 (16.9%) patients converted to glaucoma during follow-up. Those that converted were 4 years older on average and had slightly higher PSD (Table 1). Proportions converted ranged from 11.7% (hospital 9) to 20.7% (hospital 1). Distributions of other predictors were similar across groups. Proportions treated at baseline ranged from 22.0% (hospital 8) to 48.1% (hospital 10).

The discriminant power of the OHTS-EGPS model (model A) was suboptimal with a pooled c-index of 0.61 (0.60, 0.63), ranging from 0.55 to 0.67 between hospitals (Table 2). Calibration was also poor with a pooled calibration slope of 0.45 (0.38, 0.51), ranging from 0.25 to 0.64 between hospitals, where a slope of 1.00 indicates perfect calibration (Fig S3, available at www.ophthalmologyglaucoma.org/). Model performance showed no substantial differences when restricted to complete cases (Tables S3 and S4, Fig S4, available at www.ophthalmologyglaucoma.org/).

### Updating the OHTS-EGPS Model

The overall re-estimated model (model B) performed better than the OHTS-EGPS model (model A), with a pooled c-index of 0.67 (0.65–0.69) indicating better discrimination (Table 5). Calibration of the re-estimated model was good, with a pooled calibration slope of 0.96 (0.84–1.09) and good calibration across all hospitals except number 9 (Fig S5).

In both the overall and hospital-specific models (model B), re-estimated coefficients for age, VCDR, and PSD were similar to

Table 2. Performance of OHTS-EGPS Model with Original Coefficients (Model A) by Hospital

| | OHTS-EGPS (Model A) | Hospital 1 | Hospital 2 | Hospital 3 | Hospital 4 | Hospital 5 |
|---|---|---|---|---|---|---|
| n/N | | 117/564 | 64/430 | 91/628 | 111/448 | 159/1155 |
| Baseline predictor, hazard ratio (95% confidence interval) | | | | | | |
| Age (decade) | 1.26 | 1.44 (1.19–1.75) | 1.46 (1.13– 1.87) | 1.18 (0.89–1.55) | 1.20 (0.99–1.46) | 1.31 (1.09–1.56) |
| IOP (mmHg) | 1.09 | 1.00 (0.93–1.07) | 1.03 (0.94–1.13) | 0.94 (0.86–1.04) | 1.04 (0.96–1.12) | 1.04 (0.98–1.10) |
| CCT (per 40 μm thinner) | 2.04 | 0.83 (0.67–1.02) | 1.12 (0.84–1.50) | 1.20 (0.90–1.60) | 1.11 (0.89–1.39) | 1.11 (0.84–1.48) |
| VCDR (per 0.1 larger) | 1.19 | 1.03 (0.92–1.15) | 1.07 (0.92–1.25) | 1.05 (0.89–1.23) | 1.17 (1.04–1.33) | 1.17 (1.02–1.35) |
| PSD (per 0.2 dB greater) | 1.13 | 1.26 (1.13–1.39) | 1.11 (0.93–1.32) | 1.38 (1.08,–.76) | 1.25 (1.08–1.44) | 1.19 (1.05–1.34) |
| Performance measure | | | | | | |
| C-index | | 0.55 (0.50, 0.61) | 0.63 (0.56–0.71) | 0.64 (0.58–0.71) | 0.62 (0.56–0.68) | 0.64 (0.58–0.69) |
| Calibration slope | | 0.25 (0.05, 0.46) | 0.45 (0.16–0.73) | 0.55 (0.28–0.81) | 0.52 (0.30–0.74) | 0.49 (0.28–0.70) |

| | OHTS-EGPS (Model A) | Hospital 6 | Hospital 7 | Hospital 8 | Hospital 9 | Hospital 10 |
|---|---|---|---|---|---|---|
| n/N | | 512/2677 | 251/1335 | 118/876 | 82/703 | 25/214 |
| Baseline predictor, Hazard Ratio (95% Confidence Interval) | | | | | | |
| Age (decade) | 1.26 | 1.40 (1.28–1.54) | 1.23 (1.07–1.42) | 1.32 (1.09–1.61) | 1.35 (1.04–1.75) | 1.16 (0.72–1.88) |
| IOP (mmHg) | 1.09 | 0.99 (0.96–1.03) | 0.99 (0.94–1.04) | 1.03 (0.97–1.10) | 0.95 (0.86–1.05) | 0.94 (0.81–1.10) |
| CCT (per 40 μm thinner) | 2.04 | 1.10 (0.97–1.25) | 0.91 (0.75–1.10) | 0.99 (0.81–1.22) | 1.17 (0.89–1.55) | 1.37 (0.82–2.27) |
| VCDR (per 0.1 larger) | 1.19 | 1.04 (0.97–1.12) | 1.22 (1.07–1.38) | 1.00 (0.88–1.13) | 1.07 (0.63–1.84) | 1.24 (0.90–1.71) |
| PSD (per 0.2 dB greater) | 1.13 | 1.19 (1.13–1.26) | 1.30 (1.19–1.42) | 1.30 (1.16–1.46) | 1.63 (1.42–1.87) | 1.16 (0.73–1.84) |
| Performance measure | | | | | | |
| C-index | | 0.61 (0.58–0.63) | 0.62 (0.58–0.66) | 0.59 (0.54–0.64) | 0.64 (0.57–0.70) | 0.67 (0.54–0.79) |
| Calibration slope | | 0.45 (0.34–0.57) | 0.47 (0.29–0.65) | 0.36 (0.16–0.57) | 0.64 (0.29–0.98) | 0.59 (0.05–1.13) |
| Pooled c-index* = 0.61 (0.60–0.63) | | | | | | |
| Pooled calibration slope* = 0.45 (0.38–0.51) | | | | | | |

CCT = central corneal thickness; IOP = intraocular pressure; PSD = pattern standard deviation; SD = standard deviation; VCDR = vertical cup-to-disc ratio.
Multiple imputation. Re-estimated coefficients for each hospital given.
*Pooled using meta-analysis.

Table 5. Internal/External Validation of OHTS-EGPS Model with Re-estimated Coefficients (Model B) and Risk at 5 Years, by Hospital—Model Fitted in Nine Hospitals and Evaluated Separately in the Tenth Hospital

| | Imputed Dataset* n/N | C-index | Calibration Slope |
|---|---|---|---|
| Hospital | | | |
| 1 | 117/564 | 0.68 (0.62–0.74) | 0.91 (0.63–1.19) |
| 2 | 64/430 | 0.66 (0.59–0.74) | 0.75 (0.31–1.19) |
| 3 | 91/628 | 0.69 (0.61–0.76) | 1.12 (0.64–1.60) |
| 4 | 111/448 | 0.65 (0.59–0.71) | 0.91 (0.55–1.26) |
| 5 | 159/1155 | 0.66 (0.61–0.71) | 0.94 (0.66–1.22) |
| 6 | 512/2677 | 0.66 (0.63–0.68) | 0.83 (0.69–0.97) |
| 7 | 251/1335 | 0.69 (0.65–0.73) | 1.08 (0.84–1.31) |
| 8 | 118/876 | 0.68 (0.62–0.74) | 0.94 (0.62–1.26) |
| 9 | 82/703 | 0.75 (0.68–0.81) | 1.76 (1.22–2.31) |
| 10 | 25/214 | 0.67 (0.51–0.84) | 0.91 (–0.03, 1.85) |
| Pooled† | | 0.67 (0.65–0.69) | 0.96 (0.84–1.09) |

Imputed dataset.
*Model = 1-0.786^exp((0.272*(t_newage-6.262)) + (–0.006*(meaniop-24.731)) + (0.059*(t_meancct+14.098)) + (0.233*(t_meanpsd-8.379)) + (0.100*(t_meanvcdr-4.782)).
†Pooled using meta-analysis.

those in the original OHTS-EGPS model (Tables 2 and 4). In contrast, the first measurement of IOP in OHT patients was not associated with conversion risk in our dataset (hazard ratio 0.99 per unit increase [0.98–1.01] vs. 1.09 in OHTS-EGPS). The coefficient for CCT was substantially lower in our cohort (hazard ratios in re-estimated model 1.06 [0.99–1.14] vs. 2.04 in OHTS-EGPS per reduction in 40 units), indicating that a low CCT was not associated with increased glaucoma risk in our cohort.

Varying the choice of predictors in the risk prediction model had little influence on model performance (Table 6). Model coefficients for each predictor varied little among the model variants, indicating that they were largely unaffected by the addition or removal of other predictors. However, the coefficients in each model variant provide additional insight into the factors associated with conversion from OHT to glaucoma (Table 6). Sex and family history of glaucoma were not associated with conversion risk, whereas hypertension was associated with reduced risk of conversion (hazard ratio 0.81 [0.69–0.96]), and diabetes was associated with increased risk (1.27 [1.11–1.45]). Averaging IOP measurements across eyes at baseline in the main model had no influence on model performance; estimates were the same when worse eye IOP was used instead. Restricting analysis to only those with IOP ≤ 23 mmHg had similarly little influence. Restricting analysis to

Table 6. Performance of Variants of the Re-estimated OHTS-EGPS Model

| | Main Model (Model B) | Main Model and Hypertension, Family History, Diabetes, and Gender (Model C) | Worst IOP Model (Model D) | Main Model (Including ≤ 23) (Model E) |
|---|---|---|---|---|
| n/N | 1530/9030 | 1530/9030 | 1530/9030 | 362/2127 |
| Baseline predictor, Hazard Ratio (95% Confidence Interval) | | | | |
| Age, decade | 1.31 (1.24−1.38) | 1.33 (1.26−1.40) | 1.31 (1.24−1.38) | 1.28 (1.15−1.42) |
| IOP, mmHg | 0.99 (0.98−1.01) | 0.99 (0.97−1.01) | | 0.88 (0.74, 1.04) |
| CCT, per 40 μm thinner | 1.06 (0.99−1.14) | 1.06 (0.99−1.13) | 1.06 (0.99−1.14) | 1.12 (0.97−1.28) |
| PSD, per 0.2 dB greater | 1.26 (1.22−1.30) | 1.26 (1.22−1.30) | 1.26 (1.22−1.30) | 1.27 (1.20−1.35) |
| VCDR, per 0.1 larger | 1.10 (1.07−1.14) | 1.10 (1.07−1.14) | 1.10 (1.07−1.14) | 1.13 (1.05−1.21) |
| Hypertension | | 0.81 (0.69−0.96) | | |
| Family history | | 0.97 (0.86−1.10) | | |
| Diabetes | | 1.27 (1.11−1.45) | | |
| Gender | | 0.96 (0.87−1.07) | | |
| Worse IOP, mmHg | | | 0.99 (0.97−1.01) | |
| Performance measure | | | | |
| C-index* | 0.67 (0.66−0.69) | 0.68 (0.66−0.69) | 0.67 (0.66−0.69) | 0.68 (0.65−0.71) |
| Calibration slope* | 0.97 (0.87−1.08) | 0.97 (0.87−1.07) | 0.97 (0.86−1.08) | 0.97 (0.75−1.19) |

CCT = central corneal thickness; IOP = intraocular pressure; PSD = pattern standard deviation; VCDR = vertical cup-to-disc ratio.
Multiple imputation.
*Calculated within each hospital and pooled across hospitals using meta-analysis.

those not treated at VF baseline (model F, Table S7, available at www.ophthalmologyglaucoma.org/) and those never treated (model G) also made little difference, although the overall risk among those never treated was much smaller (7.3% converted in complete case analysis). These results indicate that model performance was largely unaffected by IOP-lowering treatment.

There were no substantial differences between the multiple imputation and the complete case analyses for the different model variants (Tables S8 and S7, available at www.ophthalmology glaucoma.org/).

## Influence of Baseline Variables on Treatment

Age and family history of glaucoma were not associated with treatment at VF baseline ($P > 0.05$). There was a positive association between IOP and treatment probability (odds ratio = 1.13 [1.11−1.15], $P < 0.001$), and those with thinner CCT had an increased probability of receiving treatment (odds ratio = 1.39 [1.31−1.49], $P < 0.001$).

## Discussion

We validated the OHTS-EGPS risk prediction model (model A) using a clinically-based dataset 7 times larger than the US-European dataset used for model development and 30 to 50 times larger than 4 cohorts previously used to validate the model.[14] The principal finding is that, when applied in its original form, the prediction model performed poorly. Discrimination was lower in this clinical dataset (c-index = 0.61) than reported when the OHTS-EGPS model was developed (c-index = 0.74) and in earlier validations (c-index = 0.70−0.83). In calibration terms, the model underestimated risk in all but the highest risk quintile. Reduced performance during validation is common among risk prediction models and may reflect overfitting during model development or measurement error.[18,19] However, the model itself is relatively simple—a linear combination of relevant variables and their coefficients—so overfitting is unlikely. In this study, measurement error is most likely to stem from missing data, but our results varied little when complete case analysis was performed.

A more likely explanation for the suboptimal model performance is differences in patient characteristics, disease incidence, and patient management between the populations of the original OHTS-EGPS trials and our study. As randomized clinical trials, OHTS and EGPS scheduled study visits every 6 months for 5 years,[10] whereas, in our clinical data, intervals between assessment were longer and more variable. This reflects clinical review of patients and is more representative of health care systems than data collected from RCTs. Other differences in study design were in the definition of glaucoma conversion; OHTS-EGPS used assessment of both optic disc deterioration (2 sets of photographs) and VF changes (3 consecutive abnormal tests interpreted by a reading center) to indicate conversion, whereas we used the GHT only (2 tests) without investigators' confirmation.

Across our entire dataset, the cumulative risk of conversion from OHT to glaucoma at 5 years (16.9%) was higher than the OHTS cohort but similar to the risks reported in the original EGPS study (16.8% in placebo group)[10] and in a more recent but smaller clinically-based study drawing data from 5 hospitals in England (17.5%).[12] However, there was considerable variation in conversion risk among hospitals, which may reflect differences in populations or treatment approaches among different ophthalmologists, and both discrimination and calibration were worse in hospitals with particularly high-risk patients, highlighting the sensitivity of these models to disease incidence.

Our second aim was to improve prediction by updating the model (model B), achieving modest improvements in discrimination (c-index increased to 0.67) and approaching the performance of the OHTS-EGPS model across the populations in which it was developed. These improvements are not unexpected given the internal-external validation process and the model being fitted and validated on different sections of the same base dataset. This level of discrimination is similar to that reported for validation of risk prediction tools used in other clinical areas; C-indexes of stroke risk prediction tools among women ranged from 0.61 to 0.65 for the widely used $CHADS_2$ and QStroke scores, respectively (0.63 to 0.71 among men).[20] Hepatocellular carcinoma risk models produced C-indexes ranging from 0.56 to 0.77, also displaying substantial variation in performance depending on the validation set used (e.g., the age-male-sex-ALBI-platelet count score [aMAP] model achieved 0.77 in one dataset but only 0.70 in another).[21]

The updated model was similar to OHTS-EGPS except that IOP and CCT at OHT diagnosis were not associated with glaucoma conversion risk in our dataset. Furthermore, risk factor estimates for IOP and CCT were unchanged when the model was restricted to those that were not treated at baseline or those that were never treated. It is likely that, in our study, these associations were absent because these 2 measurements strongly influence the clinician's decision whether and when to start treatment, which may in turn influence conversion risk. This hypothesis is supported by our finding that higher IOP and lower CCT were associated with higher probability of receiving treatment at baseline. Family history of glaucoma was not associated with probability of treatment, perhaps indicating that this information is not used as frequently in clinics as the more immediate IOP and CCT measurements. Our finding that IOP was not associated with risk has been observed in similar studies using UK electronic medical records[12] and clinical trial cohorts.[14]

Diabetes at baseline was associated with increased risk of glaucoma conversion, perhaps due to VF defects induced by diabetic retinopathy. Proliferative retinopathy or diabetic macular edema would likely result in abnormal glaucoma hemifield tests, triggering a conversion event. This explanation appears likely given a recent review and meta-analysis that suggested diabetes is associated with elevated IOP but not necessarily with glaucoma.[22] Systemic hypertension is associated with increased risk of glaucoma,[22] but we found a contrary result that those with hypertension were less likely to convert from OHT to glaucoma. The reduction in risk may be attributable in part to treatment of hypertension with oral beta blockers.[23]

## Strengths and Limitations

The main strength of this study was the large dataset representative of the OHT/glaucoma population in England, capturing substantial variability across the 10 sites in patient demographics, case mix, and management pathways. A key indicator was the 2-fold variability among sites in the proportion of patients treated at baseline. Although this was clinically-based data and contained missing measurements, model performance was largely unaffected. Also, measurement intervals for IOP and VF were irregular. Intraocular pressure is prone to both high short-term variability and measurement error, and, for some patients in our dataset, there was a delay between the baseline IOP measurement and the first VF assessment, so the "baseline" IOP may not represent the actual value at the start of follow-up.

A major difference between our study and OHTS was that in OHTS the majority of conversion events were determined anatomically based on analysis of sequential optic disc images by trained readers.[24] The poor performance of the OHTS-EGPS model in this study may be partially explained by our reliance on a VF-based conversion event definition. This was a pragmatic decision as VCDR measurements available in the EMR data were recorded across multiple clinics and so were more liable to interobserver variability than those in OHTS-EGPS.

Also, there was only a single cup-to-disc ratio measurement for each eye for each visit in the EMR (Fig S1). We have referred to this as VCDR as the vertical measurement is most commonly taken in these clinics, but it is possible that some entries may contain different cup-to-disc ratio (CDR) measurements. Despite this uncertainty, the CDR measurements in our study show conceptual validity in that we found consistent positive associations between them and increased glaucoma risk in all but one of the risk models. Furthermore, the Heidelberg retinal tomography or OCT images used to make these measurements were not available in our data; otherwise, we would have considered imaging-derived measurements and outcomes.

In OHTS, an endpoint committee determined whether conversion to glaucoma had occurred, accounting for the presence of ocular comorbidities that may have induced VF defects (e.g., age-related macular degeneration, retinal vein occlusion). We attempted to disentangle glaucoma conversion by excluding eyes with these conditions at baseline and by censoring eyes that developed them during follow-up at the date of comorbidity diagnosis. Thus, our estimates of conversion risk are likely to have been independent of these conditions, but we were also unable to explore the possible influence of comorbidities on glaucoma conversion risk (e.g., does retinal vein occlusion [RVO] increase risk of conversion?).

## Further Work

Our current model uses only baseline data. A possible extension would be to use measurements from the first 2 or 3 clinic visits to capture initial responses to treatment and improve model performance. Survival analysis specifying IOP, medication status, and visual field parameters as time-varying covariates would be one approach. Given the variability among individuals in monitoring intervals and likely responses, more flexible models fitted using machine learning could also be considered. Finally, this dataset is a valuable resource to investigate prediction models for glaucoma progression, including both those that converted from OHT used in this analysis and those with pre-existing glaucoma.

## Conclusions

We validated the OHTS-EGPS risk prediction model for conversion from OHT to glaucoma using electronic data from a large cohort. By refitting the model, we achieved modest improvements in model performance, warranting further research on how these predictions might be incorporated into clinical practice.

## Acknowledgments

The authors wish to thank leads at all study sites: Moorfields Eye Hospital NHS Foundation Trust, Nottingham University Hospitals NHS Trust, University Hospitals Bristol and Weston NHS Foundation Trust, Mid-Yorkshire Hospitals NHS Trust, University Hospitals Plymouth NHS Trust, Gloucestershire Hospitals NHS Foundation Trust, Imperial College Healthcare NHS Trust, East Suffolk and North Essex NHS Foundation Trust, Wirral University Teaching Hospital NHS Foundation Trust, Bedford Hospital NHS Trust, and Calderdale and Huddersfield NHS Foundation Trust.

## Footnotes and Disclosures

Originally received: February 28, 2024.
Final revision: October 25, 2024.
Accepted: October 28, 2024.
Available online: November 4, 2024. Manuscript no. OGLA-D-24-00076.

[1] Centre for Public Health, Queen's University Belfast, Belfast, UK.

[2] Belfast Health and Social Care Trust, Belfast, UK.

[3] Optometry and Visual Sciences, City, University of London, London, UK.

[4] Moorfields Eye Hospital NHS Foundation Trust, London, UK.

[5] Nottingham University Hospitals NHS Trust, Nottingham, UK.

[6] Health Economics Research Unit, University of Aberdeen, Aberdeen, UK.

[7] School of Optometry and Visual Sciences, Cardiff University, Cardiff, UK.

[8] Institute of Applied Health Research, University of Birmingham, Birmingham, UK.

*Joint lead authors.

Disclosures:

All authors have completed and submitted the ICMJE disclosures form.

The authors made the following disclosures:

D.M.W.: Grants − National Institute of Health Research.

C.C.: Grants − National Institute of Health Research.

D.P.C.: Consultant − Apellis; Grants − Apellis, Santen; Honoraria − Allergan/Abbvie, Janssen, Santen, Thea.

G.G.: Consultant − Alcon, Allergan, Belkin, Elios, Equinox, Genentech/Roche, Glaukos, Ivantis, McKinsey, Rayner, Reichert, Ripple Therapeutics, Santen, Sight Sciences, Thea, Vialase, Visufarma, Zeiss; Grants − Thea, Santen; Honoraria − Alcon, Allergan, Belkin, Glaukos, Ivantis, Lumibird, McKinsey, Reichert, Sight Sciences, Thea; Travel expenses − Ivantis, Thea; Board membership − Glaucoma UK, UK & Ireland Glaucoma Society.

A.J.K.: Consultant − Thea, Abbvie.

R.H.: Grants − National Institute of Health Research.

Y.T.: Grants − National Institute of Health Research.

Supported by the National Institute for Health and Care Research—Health Technology Assessment—Glaucoma Risk Prediction in Ocular Hypertension (GRIP) (NIHR131808).

Giovanni Montesano, PhD, an editor of this journal, was recused from the peer-review process of this article and had no access to information regarding its peer-review.

HUMAN SUBJECTS: Human subjects were included in this study. This study adheres to the tenets of the Declaration of Helsinki. Ethical approval for use of these data was obtained (REC reference 21/EE/0109) and permissions received from each center. As all records were anonymized prior to extraction, patient consent was not required.

No animal subjects were included in this study.

Author Contributions:

Conception and design: Wright, Azuara-Blanco, Cardwell, Montesano, Crabb, Gazzard, King, Hernández, Morgan, Higgins, Takwoingi

Data collection: Wright, Azuara-Blanco, Cardwell, Montesano, Crabb, Gazzard, King

Analysis and interpretation: Wright, Azuara-Blanco, Cardwell, Montesano, Crabb, Gazzard, King, Hernández, Morgan, Higgins, Takwoingi

Obtained funding: Azuara-Blanco, Cardwell, Crabb, Gazzard, King, Hernandez, Morgan, Takwoingi

Overall responsibility: Wright, Azuara-Blanco, Cardwell, Montesano, Crabb, Gazzard, King, Hernández, Morgan, Higgins, Takwoingi

Abbreviations and Acronyms:
**CCT** = central corneal thickness; **EMR** = electronic medical record; **GHT** = Glaucoma Hemifield Test; **IOP** = intraocular pressure; **OHT** = ocular hypertension; **PSD** = pattern standard deviation; **VCDR** = vertical cup-to-disc ratio; **VF** = visual field.

Keywords:
Ocular hypertension, Glaucoma, Risk prediction, Electronic medical records.

Correspondence:
David M. Wright, Institute of Clinical Sciences Block A, Royal Victoria Hospital, Grosvenor Road, Belfast BT12 6BA, UK. E-mail: d.wright@qub.ac.uk.

## References

1. Bourne RRA, Jonas JB, Bron AM, et al. Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe in 2015: magnitude, temporal trends and projections. *Br J Ophthalmol*. 2018;102:575−585.
2. Adelson JD, Bourne RRA, Briant PS, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study. *Lancet Global Health*. 2020. https://www.thelancet.com/journals/langlo/article/PIIS2214-109X(20)30489-7/abstract. Accessed December 7, 2020.
3. Tham Y-C, Li X, Wong TY, et al. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*. 2014;121:2081−2090.
4. Gordon MO, Beiser JA, Brandt JD, et al. The ocular hypertension treatment study: baseline factors that predict the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120:714−720.

5. Nislawati R, Zainal ATF, Ismail A, et al. Role of hypertension as a risk factor for open-angle glaucoma: a systematic review and meta-analysis. *BMJ Open Ophthalmology*. 2021;6, e000798.
6. Weinreb RN, Aung T, Medeiros FA. The pathophysiology and treatment of glaucoma: a review. *JAMA*. 2014;311:1901−1911.
7. Suzuki Y, Iwase A, Araie M, et al. Risk factors for open-angle glaucoma in a Japanese population: the tajimi study. *Ophthalmology*. 2006;113:1613−1617.
8. Royal College of Ophthalmologists. The way forward - glaucoma, London. https://www.rcophth.ac.uk/wp-content/uploads/2015/10/RCOphth-The-Way-Forward-Glaucoma-300117.pdf; 2017. Accessed November 9, 2020.
9. NICE. Glaucoma: diagnosis and management. https://www.nice.org.uk/guidance/ng81; 2017. Accessed July 25, 2023.
10. Ocular Hypertension Treatment Study Group, European Glaucoma Prevention Study Group. Validated prediction model for the development of primary open-angle glaucoma in individuals with ocular hypertension. *Ophthalmology*. 2007;114:10−19.e2.
11. Yohannan J, Wang J, Brown J, et al. Evidence-based criteria for assessment of visual field reliability. *Ophthalmology*. 2017;124:1612−1620.
12. Kelly SR, Khawaja AP, Bryan SR, et al. Progression from ocular hypertension to visual field loss in the English hospital eye service. *Br J Ophthalmol*. 2020;104:1406−1411.
13. Anon. Research Registry - ID: researchregistry6647. Research Registry. https://www.researchregistry.com/browse-the-registry Accessed July 25, 2023.
14. Burr J, Botello-Pinzon P, Takwoingi Y, et al. Surveillance for ocular hypertension: an evidence synthesis and economic evaluation. *Health Technol Assess*. 2012;16. https://www.journalslibrary.nihr.ac.uk/hta/hta16290/. Accessed November 21, 2022.
15. Snell KI, Ensor J, Debray TP, et al. Meta-analysis of prediction model performance across multiple studies: which scale helps ensure between-study normality for the *C* -statistic and calibration measures? *Stat Methods Med Res*. 2018;27:3505−3522.
16. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statist Med*. 2011;30:377−399.
17. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338, b2393.
18. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21: 128−138.
19. Van Calster B, McLernon DJ, van Smeden M, et al. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17:230.
20. Hippisley-Cox J, Coupland C, Brindle P. Derivation and validation of QStroke score for predicting risk of ischaemic stroke in primary care and comparison with other risk scores: a prospective open cohort study. *BMJ*. 2013;346, f2573.
21. Innes H, Jepsen P, McDonald S, et al. Performance of models to predict hepatocellular carcinoma risk among UK patients with cirrhosis and cured HCV infection. *JHEP Reports*. 2021;3, 100384.
22. Grzybowski A, Och M, Kanclerz P, et al. Primary open angle glaucoma and vascular risk factors: a review of population based studies from 1990 to 2019. *J Clin Med*. 2020;9:761.
23. Langman MJS, Lancashire RJ, Cheng KK, Stewart PM. Systemic hypertension and glaucoma: mechanisms in common and co-occurrence. *Br J Ophthalmol*. 2005;89: 960−963.
24. Kass MA, Heuer DK, Higginbotham EJ, et al. The ocular hypertension treatment study: a randomized trial determines that topical ocular hypotensive medication delays or prevents the onset of primary open-angle glaucoma. *Arch Ophthalmol*. 2002;120:701−713.