Personalized Model to Predict Keratoconus Progression From Demographic, Topographic, and Genetic Data

HOWARD P. MAILE, JI-PENG OLIVIA LI, MARY D. FORTUNE, PATRICK ROYSTON, MARCELLO T. LEUCCI, ISMAIL MOGHUL, ANITA SZABO, KONSTANTINOS BALASKAS, BRUCE D. ALLAN, ALISON J. HARDCASTLE, PIRRO HYSI, NIKOLAS PONTIKOS, STEPHEN J. TUFT, AND DANIEL M. GORE

• PURPOSE: To generate a prognostic model to predict keratoconus progression to corneal crosslinking (CXL).
• DESIGN: Retrospective cohort study.
• METHODS: We recruited 5025 patients (9341 eyes) with early keratoconus between January 2011 and November 2020. Genetic data from 926 patients were available. We investigated both keratometry or CXL as end points for progression and used the Royston-Parmar method on the proportional hazards scale to generate a prognostic model. We calculated hazard ratios (HRs) for each significant covariate, with explained variation and discrimination, and performed internal-external cross validation by geographic regions.
• RESULTS: After exclusions, model fitting comprised 8701 eyes, of which 3232 underwent CXL. For early keratoconus, CXL provided a more robust prognostic model than keratometric progression. The final model explained 33% of the variation in time to event: age HR (95% CI) 0.9 (0.90-0.91), maximum anterior keratometry 1.08 (1.07-1.09), and minimum corneal thickness 0.95 (0.93-0.96) as significant covariates. Single-nucleotide polymorphisms (SNPs) associated with keratoconus (n=28) did not significantly contribute to the model. The predicted time-to-event curves closely followed the observed curves during internal-external validation. Differences in discrimination between geographic regions was low, suggesting the model maintained its predictive ability.
• CONCLUSIONS: A prognostic model to predict keratoconus progression could aid patient empowerment, triage, and service provision. Age at presentation is the most significant predictor of progression risk. Candidate SNPs associated with keratoconus do not contribute to progression risk. (Am J Ophthalmol 2022;240: 321–329. © 2022 Elsevier Inc. All rights reserved.)

Keratoconus is a common corneal ectasia that causes irregular astigmatism, scarring, and loss of vision. Thinning and steepening can progress through childhood and early adulthood, but the shape of most eyes stabilizes by the third or fourth decade. Without intervention, keratoconus can lead to severe visual loss, with approximately 10% of eyes eventually requiring corneal transplantation.[1] Corneal crosslinking (CXL) by topical application of riboflavin, followed by irradiation with UV-A light, can arrest progression of keratoconus in up to 88% to 100% of eyes even when there is relatively advanced disease.[2–6] The potential benefit of CXL is to prevent visual deterioration with a relatively low risk procedure that is cost effective for health care providers.[7–9]

However, CXL is usually not offered to all patients at presentation because the disease may have already stabilized. In the recent KERALINK study, 43% of children aged <17 years at presentation had not progressed after 18 months.[10] The definition of progression also varies with the severity of keratoconus, but for early disease a common threshold is either an increase in the maximum keratometry (Kmax) of >1 diopter (D), a change in the manifest refractive spherical equivalent of >0.50 D, or an increase in manifest refractive cylinder of >1 D.[2,11] Depending on the rate of progression this threshold may be passed in a few months, years, or not at all. At the first assessment it can be a challenge to distinguish eyes that are at risk of rapid progression from those where it is safe to monitor. Unnecessary review visits are a burden to the patient and the care system.

We considered the date of numeric progression,[6] as well as the date when CXL was performed, as alternative end points to define keratoconus progression. Although the use of keratometry as an end point may appear the more objective method, there is variability on the definition of progression reported in the literature and conclusions may vary with the definition that is adopted.[11–14] Repeatability thresholds are not usually tailored to individual eyes (ie, an increase in Kmax by 1 D is not significant in all eyes) although there is growing evidence on the variability of mea-

AJO.com Supplemental Material available at AJO.com.
Accepted for publication April 13, 2022.

From the UCL Institute of Ophthalmology (H.P.M., I.M., A.S., A.J.H., N.P.), London; Moorfields Eye Hospital NHS Foundation Trust (J.-P.O.L., M.T.L., K.B., B.D.A., S.J.T., D.M.G.), London; MRC Biostatistics Unit, Cambridge Institute of Public Health, University of Cambridge (M.D.F.), Cambridge; MRC Clinical Trials Unit at UCL (P.R.), London; Section of Ophthalmology, School of Life Course Sciences, King's College London (P.H.), London; Department of Twin Research and Genetic Epidemiology, King's College London (P.H.), United Kingdom

Inquiries to Daniel M. Gore, Moorfields Eye Hospital, London, United Kingdom; e-mail: daniel.gore1@nhs.net

0002-9394/$36.00
https://doi.org/10.1016/j.ajo.2022.04.004

surements in more advanced disease and the need for tailoring numerical progression definitions to the disease state, and distinguishing real progression from inherent variability of measurement modalities.[15–17]

Finally, patients who receive CXL prior to progression must be censored from the data set even though these eyes are likely to have been at risk of progression. This type of informative censoring creates a bias.[18] In contrast, the time to CXL depends on several variables that include numeric disease progression, but also incorporates patient-specific risk factors for future progression. Its strength is that it is an easily comprehensible and meaningful end point for patients. It encompasses individual risk factors that are not considered when imaging is used in isolation and it has been used by others as defining the event of interest.[19]

For these reasons, we have used demographic and serial tomography data from a large cohort of patients to generate a time-to-event model to predict the probability of an individual progressing to CXL. Because the Cox proportional hazards method does not generate smooth time-to-event curves, we used the Royston-Parmar model to achieve direction estimates of the hazard function.[20] We also performed a further analysis of a subset of patients who had genetic data in the form of single-nucleotide polymorphisms (SNPs) generated as part of a study to determine keratoconus risk.[21]

METHODS

• **COHORT:** The study protocol was reviewed and approved by the Clinical Audit Assessment Committee of Moorfields Eye Hospital NHS Foundation Trust (reference CA17/CED/03). Institutional Review Board (IRB) approval was obtained, and individual patient consent was not required. The study conformed to the tenets of the Declaration of Helsinki. We identified from the Moorfields Eye Hospital electronic health record database (OpenEyes) patients aged ≥13 years diagnosed with clinical or suspected keratoconus who attended our Early Keratoconus Clinic between January 2011 and November 2020.

Clinical data included keratometry (Kmax, Front K1, Front K2, Back K1, Back K2), and pachymetry (minimum corneal thickness) captured by Scheimpflug tomography (Pentacam HR; Oculus GmbH). We only included scans with a quality score of "good" or "OK," and where multiple scans were taken on the same day, we used the mean value. The date of all CXL procedures was recorded.

The protocol for offering CXL throughout the study period was as follows: (1) a documented history prior to referral to the Early Keratoconus Clinic of our hospital of significant recent disease progression,[6] (2) a change in contemporary measurements of 95% above the repeatability limits of the baseline measurements as shown in Supplementary Table S1 (available at http://www.ajo.com),[6] or (3) a patient considered by a clinician to be at high risk of progression despite their not fulfilling the above 2 criteria. Exclusion criteria included pregnancy or breastfeeding, uncontrolled ocular surface disease, or a minimum corneal thickness less than 375 μm.

All the data used for model fitting started from the first appointment in the Early Keratoconus Clinic. Patient demographics included age, gender, smoking status (current or ex-/nonsmoker), ethnicity, and postcode. Ethnicity was coded as 1 for "Black" or "South Asian or South Asian British" and 0 for any other category (excluding missing values). Before model fitting, the pachymetry in micrometers was divided by 10 to generate a meaningful scale. For the primary analysis, eyes with any missing data were excluded. We also explored multiple imputation, which avoids data exclusion by generating multiple versions of the data set, with missing values replaced with values sampled from an appropriate distribution.

To see whether genetic data can help predict keratoconus progression, we used 28 candidate SNPs from a recent keratoconus genome-wide association study that contained 926 patients from Moorfields Eye Hospital.[21] The SNP data were encoded as 0 (homozygous reference genotype), 1 (heterozygous genotype), or 2 (homozygous variant genotype). We chose to use an additive encoding; thus, the risk of disease increases additively with the degree of genetic variation.[22] Anonymized data were then exported to Excel software for analysis (version 15.24 2016, Microsoft Corp).

• **MODEL FITTING AND COVARIATE SELECTION:** A Royston-Parmar flexible parametric survival model was fitted to the data to predict the probability of an eye progressing to CXL.[23] Initial analysis of the covariates was performed by univariate analysis using the same model characteristics as the multivariable model. When selecting covariates for the final multivariable model, we used backwards stepwise selection with a significance level of 0.05. We used linear covariates for ease of interpretation of our final model. To create a more parsimonious model we examined the effect on explained variation and discrimination of removing single variables from the model.

• **KERATOMETRIC PROGRESSION SENSITIVITY ANALYSIS:** We included a sensitivity analysis in which we investigated keratometric progression as an alternative end point. Keratometric progression was defined using thresholds from Gore and associates.[6] When using numerical thresholds to define progression, the appointments for eyes beyond the date of CXL cannot be used. However, censoring these eyes at the date of CXL represents informative censoring. Based on the recommendations of Clarke and associates[18] for investigating the impact of informative censoring, we generated a "best case" data set where eyes were censored at the CXL date and a "worst case" data set where patients were assumed to progress at the CXL date.

The corresponding Kaplan-Meier curves were plotted to provide a visual comparison of the 2 data sets. A Royston-Parmar model was then fitted on both data sets. We used the same techniques (backward stepwise selection, significance level of .05) as described in the previous section to fit the model and compare the explained variation and hazard ratios.

• **MULTIVARIABLE MODEL VALIDATION:** We validated the model using internal-external cross validation in which we split the data set by geographical region.[24,25] For the *k*th region, the model is fitted on the full data set excluding region *k*, and then Kaplan-Meier curves and predicted survival curves were generated for region *k*. Seven geographical regions were created based on the patient's postcode as shown in Supplementary Figure S1 (available at http://www.ajo.com).

To quantitatively assess the validation, Royston and Sauerbrei's *D* statistic was calculated for both the model fitted from data excluding region *k* ($D_{(k)}$) and also the model applied to region *k* (D_k).[26] The difference between these 2 discrimination metrics ($D_k - D_{(k)}$) was calculated with its corresponding standard error to assess the predictive ability of the model. To demonstrate how the model could be used in practice, we include 3 hypothetical patients' eyes with different progression risk profiles (high, medium, low risk) and plot the predicted time-to-event curve for each shown in Figure 2.

• **STATISTICAL ANALYSIS:** The event of interest was defined as the date that the eye underwent CXL. We calculated the time-to-event as the difference between the first appointment in our service and the date of CXL (or the last patient appointment in the case of censoring). Because we had paired observations (eyes), we used variance-corrected models to account for correlation between eyes and to ensure that robust SEs were produced. The choice of scale and selection of degrees of freedom for the Royston-Parmar model was informed by inspecting the Akaike information criterion and Bayes information criterion,[20] and the results of this were balanced with ease of interpretation. See Supplementary Table S2 and Supplementary Material S1 (available at http://www.ajo.com) for further explanation.

Royston and Sauerbrei's *D* statistic was used as a measure of discrimination and R^2_D as a measure of explained variation (both calculated on the natural scale of the model). Although all of the primary results were generated from a complete case analysis, we performed an additional analysis using multiple chained imputation (predictive mean matching approach with 5 nearest neighbors). Model fitting was performed in Stata 13 (StataCorp LP), and the Royston-Parmar model was fitted using the stpm2 package from Stata 13.

RESULTS

• **COHORT:** From a potential of 9341 eyes (4316 pairs of eyes and 709 individual eyes), the final model used 8701 eyes of 4823 patients, with 3232 eyes that had CXL. The mean age was 28.3 years with SD of 7.1 years. We excluded 640 eyes with missing data. Table 1 summarizes the available covariates along with missing data percentages. See Supplementary Material S2 and Supplementary Table S3 (available at http://www.ajo.com) for a description of the multiple imputation results.

• **MODEL FITTING AND COVARIATE SELECTION (GENETIC DATA):** We analyzed patients with genetic data separately because these data were only available for ~14% of patients. Of 926 patients (1852 eyes) with genetic data, 531 eyes were excluded with incomplete keratometry or CXL data, which left 1321 eyes, of which 665 had CXL. With univariate analysis of the 28 SNPs, only rs72631889 was found to be significant ($P = .01$) (Supplementary Table S4 [available at http://www.ajo.com]). We then produced a multivariable model via backward selection on this subset of eyes using corneal data, patient data, and rs72631889 as an additional covariate as shown in Supplementary Table S5 (available at http://www.ajo.com). However, rs72631889, although significant ($P = .005$), had a negligible contribution (0.3%) to the explained variation in the final model.

• **MODEL FITTING AND COVARIATE SELECTION (EXCLUDING GENETIC DATA):** The results of the univariate time-to-event analysis on the hazards scale using a Royston-Parmar flexible parametric model is shown in Table 2. Genetic data was excluded from this analysis. All variables except smoking status were significant. The explained variation (R^2_D) and discrimination (*D*) were highest for age (17%) and Kmax (15%) with Front K1, Front K2, Back K1, Back K2, and pachymetry each explaining 6% to 10% of the variation. Notably, gender and ethnicity, although significant in the univariate analysis, did not contribute to explained variation. The hazard ratios for significant covariates indicate that increasing age at presentation, greater pachymetry, and flatter (less negative) posterior keratometry values decrease the risk of having CXL, whereas steeper anterior keratometry values and male gender increase the risk of having CXL.

When we fitted a multivariable model, the significant covariates were age, Kmax, Front K1, Front K2, and pachymetry (Table 2). When we removed single variables from the model, the effect this had on explained variation and discrimination is shown in Supplementary Table S6 (available at http://www.ajo.com). Age was the most important covariate (16.7%), with Kmax contributing ~5% of explained variation. K1, K2 and pachymetry

TABLE 1. Summary Statistics for the Available Covariates at the First Examination for 9341 Eyes Recorded at First Visit.

Covariate	Type	Mean	SD	No. of Eyes	Missing No. (%)
Front K1 (D)	Numeric	45.31	3.86	8813	528 (5.7)
Front K2 (D)	Numeric	48.39	4.85	8839	502 (5.4)
Back K1 (D)	Numeric	–6.53	0.75	7949	1392 (14.9)
Back K2 (D)	Numeric	–7.23	0.93	8702	639 (6.8)
Kmax (D)	Numeric	54.14	8.01	8834	507 (5.4)
Pachymetry (μm)	Numeric	462.92	46.15	8946	395 (4.2)
Age (y)	Numeric	28.28	7.10	9341	0 (0)
Genetic data[a]	Ordinal	N/A	N/A	1141	8020 (85.9)
Self-reported Black or Asian ethnicity[b]	Categorical (59.9% Black or Asian)	N/A	N/A	4889	4452 (47.7)
Male gender	Categorical (67% male)	N/A	N/A	9341	0 (0)
Smoker[c]	Categorical (4.5% smoker)	N/A	N/A	9341	0 (0)

Back K1 = flattest posterior keratometry, Back K2 = steepest posterior keratometry, Front K1 = flattest anterior keratometry, Front K2 = steepest anterior keratometry, Kmax = maximum keratometry, pachymetry = minimum corneal thickness, N/A = not applicable.

[a]Genetic data composed of 28 SNPs and was encoded in an additive fashion (0, 1, 2).

[b]1 = Black or Asian, 0 = otherwise.

[c]0 = nonsmoker/ex-smoker, 1 = current smoker.

TABLE 2. Univariable and Final Multivariable Model for All Considered Covariables Excluding Genetic Data in the Training Data Set Fitted on the Hazards Scale With 5 Degrees of Freedom

	Univariable XXXX (no. of eyes = 9341)				Multivariable XXXX (no. of eyes = 8701)	
Covariate	Hazard Ratio (95% CI)	*P* Value	R^2_D (%)	*D*	Hazard Ratio (95% CI)	*P* Value
Ethnicity	1.14 (1.02, 1.27)	.02	0.4	0.13	N/A	N/A
Smoker[a]	1.07 (0.9, 1.28)	.46	0.1	0.05	N/A	N/A
Male gender	1.11 (1.01, 1.21)	.02	0.2	0.10	N/A	N/A
Age at presentation	0.91 (0.9, 0.92)	<.001	16.7	0.92	0.9 (0.90, 0.91)	<.001
Kmax	1.06 (1.05, 1.06)	<.001	14.9	0.86	1.08 (1.07, 1.09)	<.001
Front K1	1.09 (1.08, 1.1)	<.001	7.0	0.56	0.93 (0.91, 0.94)	<.001
Front K2	1.08 (1.07, 1.08)	<.001	9.8	0.67	N/A	N/A
Back K1[b]	0.67 (0.64, 0.71)	<.001	5.9	0.51	N/A	N/A
Back K2[b]	0.7 (0.67, 0.72)	<.001	8.4	0.62	N/A	N/A
Pachymetry 10[b]	0.93 (0.92, 0.94)	<.001	7.5	0.58	0.95 (0.93, 0.96)	<.001

Back K1 = flattest posterior keratometry, Back K2 = steepest posterior keratometry, D = Royston and Sauerbrei's *D* statistic (used as a measure of discrimination), Front K1 = flattest anterior keratometry, Front K2 = steepest anterior keratometry, Kmax = maximum keratometry, N/A = not applicable due to this variable not being included in the final model, pachymetry = minimum corneal thickness, R^2_D = explained variation.

[a]0 = nonsmoker/ex-smoker; 1 = current smoker.

[b]Back K1 and Back K2 are negative values such that patients with advanced keratoconus are typically associated with large negative values. A hazard ratio below 1 indicates that as measurements become more positive, the risk of progression decreases.[c]Minimum pachymetry in steps of 10 μm.

had a small effect (<1%) when removed individually. We chose a model without K2 on the basis of parsimony, which was supported by the fact that K1 and K2 were highly correlated ($R^2 = 0.91$) as shown in Supplementary Figure S2 (available at http://www.ajo.com).

The final fitted model hazard ratios can be seen on the multivariable column of Table 2. It is notable that an increase in K1 now has a protective effect in the final model. The explained variation and discrimination for the final model were 32.7% and 1.43, respectively.[27] The opposing effect of Kmax and Front K1 can be explained by examining their regression coefficients before converting to hazard ratios; Kmax has a positive coefficient (0.0795) and Front K1 has a negative coefficient (–0.0749). This is logically similar to including the combined covariate (Kmax-Front K1) in the model, which can be viewed clinically as a proxy for irregular astigmatism. We also investigated combining K1 and K2 into a single covariate as K2-K1 (standard definition of astigmatism), but the corresponding *P* value was not significant.

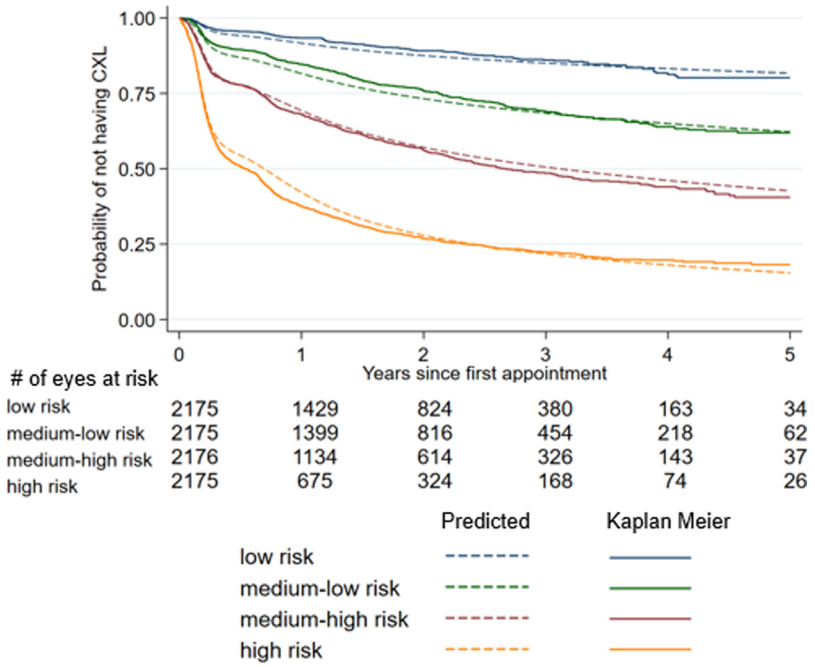

FIGURE 1. Chart showing how the Royston-Parmar model fits the entire data set. We split the eyes into 4 risk groups by their prognostic index: <25th centile (low risk), 25-50th centile (medium-low risk), 50-75th centile (medium-high risk), >75th centile (high risk). The number of eyes at risk corresponds to the Kaplan-Meier curves.

Figure 1 visually depicts the result of applying the final model to the original data set. As expected, the predicted mean survival curves closely follow the Kaplan-Meier curves. To demonstrate the use of the model in clinical practice, survival curves for 3 hypothetical patients followed for 5 years are shown in Figure 2. We have also produced a web application from the model, which can be accessed at http://beta.moorfieldscxl.com.

• **KERATOMETRIC PROGRESSION SENSITIVITY ANALYSIS:** The results of the keratometric progression sensitivity analysis can be found in the Supplementary Material. By examining the Kaplan-Meier curves in Supplementary Figure S3, we can see that the best case time-to-event curve indicates a 40% survival probability at 5 years while the worst case curve indicates a 27% survival probability at 5 years. This 13% difference in survival probability at 5 years represents the upper bound of the discrepancy in survival probability within the data. After fitting the Royston-Parmar model, among the hazard ratios that overlap (age, Kmax, K2), there was reasonable similarity (Supplementary Tables S8 and S9). Most important, the model fitted to the best case had an explained variation of 11% compared to 23% for the worst case, indicating a significant difference in model performance depending on the assumptions used for handling eyes that received CXL.

• **MULTIVARIABLE MODEL VALIDATION:** When performing validation using internal-external cross validation, Figure 3 shows the ability of our final model to predict keratoconus progression across different geographic regions. We did not identify any significant differences in prognostic factors across regions. The model prediction curves generally follow the Kaplan-Meier curves. Notably, region 5 (Southwest Greater London) and region 7 (other regions) have a worse predictive performance than the other regions, indicating that these regions have different characteristics compared with the remainder of the data set used for model fitting. This could be due to differing patient characteristics, such as complex cases that required referral to our tertiary referral center rather than being managed locally.

Overall, the prediction becomes less accurate over time, which is expected because of low numbers with follow-up beyond 3 years. Supplementary Table S7 displays quantitative validation results of the model using internal external validation. The difference column $D_k - D_{(k)}$ is a measure of predictive ability. Region 7 (other regions outside of Greater London) has the greatest discrepancy in discrimination (–0.26), which indicates that the model fitted when excluding region 7 had greater discriminative ability than when applied to region 7 alone.

DISCUSSION

In this study, we have incorporated demographic, keratometric, and genetic data to generate a prognostic model

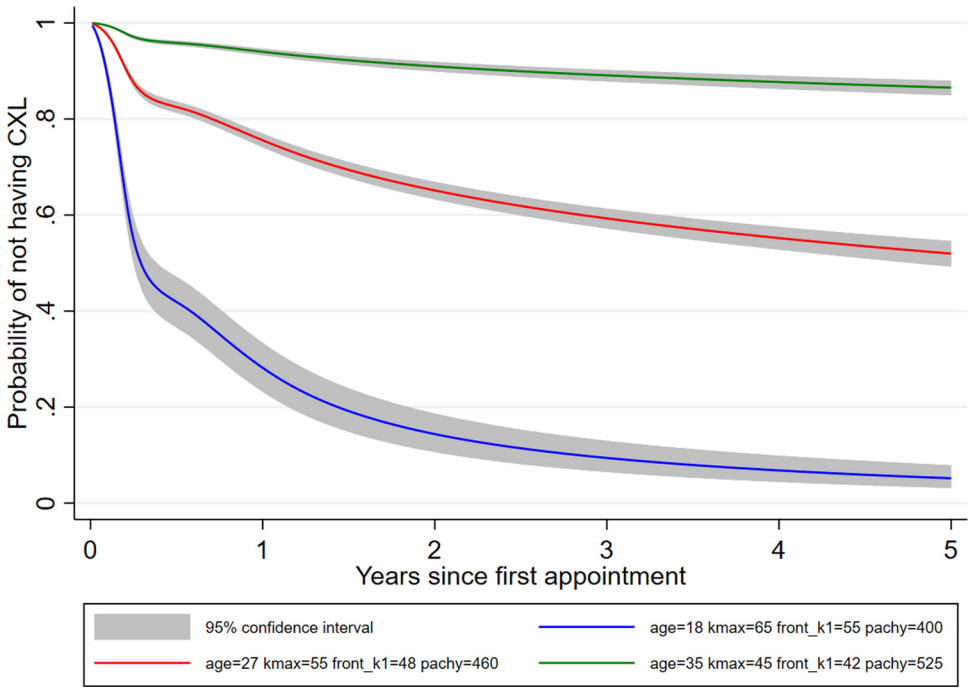

FIGURE 2. Time-to-event curves that predict the risk of progression to CXL for 3 hypothetical patient profiles. The blue line represents a high-risk patient who has a 95% probability of progressing to CXL at 5 years. The red line is a medium-risk patient who has a 48% probability of progressing to CXL at 5 years. The green line is a low-risk patient who has a 14% probability of progressing to CXL at 5 years. The equation used to generate the curves is: $S(t) = e^{-H(t)}$, where $H(t)$ is the cumulative hazard function and is commonly expressed as $\ln(H(t)) = s(ln(t)) + x\beta$, where $s(ln(t))$ is a restricted cubic spline function of log time, β is the vector of coefficients, and x is the vector of covariates. For further details of the derivation, we refer the reader to Patrick Royston.[20] pachy = pachymetry.

of keratoconus progression to CXL. We have shown that parameters recorded at the first examination (age, Kmax, Front K1, minimum pachymetry) can produce a time-to-event curve to calculate a personalized risk for keratoconus progression. Although we chose time to CXL rather than keratometric progression as the end point for the time-to-event analysis, we performed a sensitivity analysis using keratometric progression and found that a CXL model accounts for a much higher proportion of the explained variation (33%) compared with the keratometric model (11% or 23% for best and worst case, respectively).

The opposing effects of Kmax and Front K1 were unexpected, but similar to including the combined covariate (Kmax – Front K1) in the model; a possible explanation is that the opposing effect is the result of an increase in irregular astigmatism. Of the significant covariates in our model, younger age made the greatest contribution to our model. Thus, one should have a lower threshold for treatment in younger patients.

When applying internal-external cross validation, the survival curves closely followed the Kaplan-Meier survival curves for each of the geographic regions, which indicates generalizability, and model discrimination between training and cross-validation groups was similar, indicating that the predictive ability is well maintained. Finally, our SNP genetic data had limited additional predictive utility for keratoconus progression. However, the genetic data set was relatively small (926 patients), and recruitment was based on the presence of keratoconus, as opposed to the severity of keratoconus, or any other index of risk of rapid progression.

The Royston-Parmar model has previously been used to predict the likelihood of the worst eye of patients with keratoconus progressing to corneal transplantation.[28] In their final model, Quartilho and associates chose 3 significant covariates: Kmax, age, and ethnicity. The reported covariate hazard ratios that overlap with our study (Kmax and age) were different in magnitude but in the same direction. When performing internal validation, their model exhibited good predictive ability.

They produced time-dependent receiver operating characteristic curves using the validation set and found 1-year sensitivity and specificity to be 92.8% and 94.6%, respectively. Using logistic regression, Kato and associates found that the 2 strongest factors associated with the requirement for CXL were age and Kmax, which is consistent with our findings.[19] Moreover, the team went on to find that age combined with corneal tomography maps was able to predict progression and need for crosslinking using deep learning.[29]

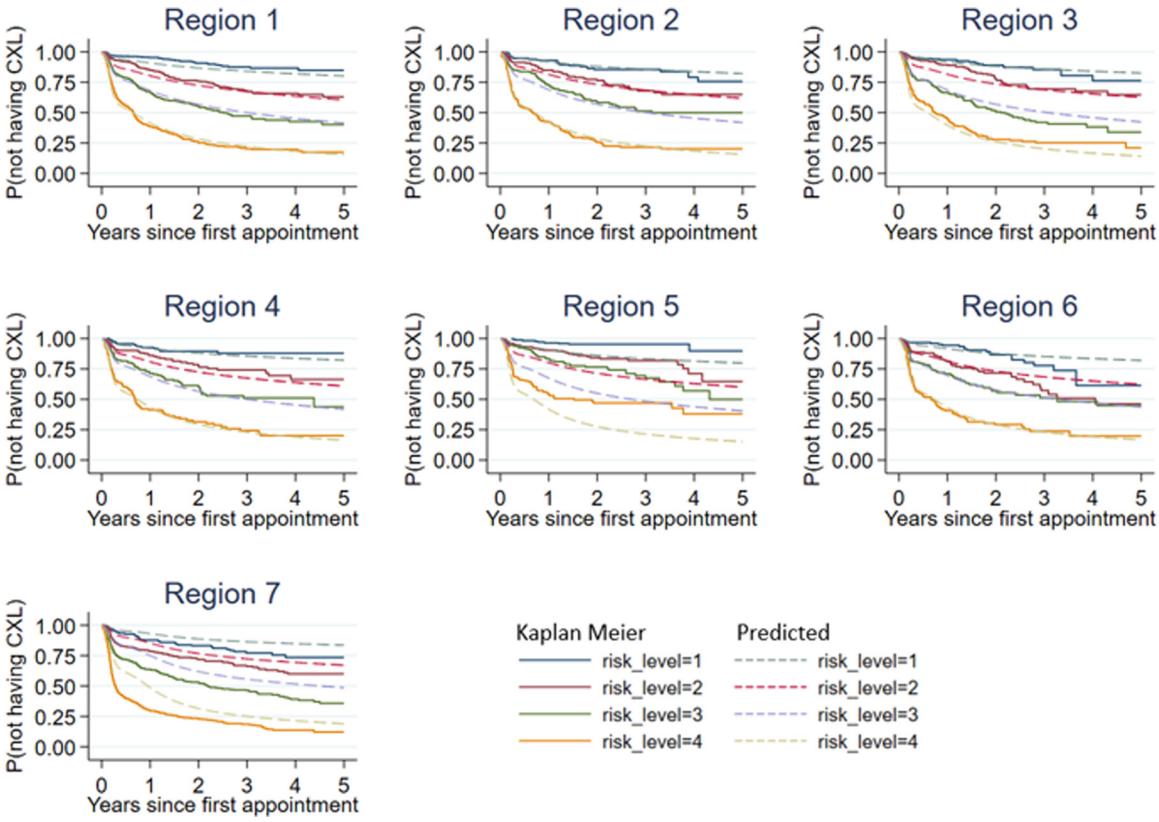

FIGURE 3. Predicted and observed survival curves for 7 postal code regions of Greater London as shown in Supplementary Figure S1 (available at http://www.ajo.com) using internal-external cross-validation. We split the eyes into 4 risk groups by their prognostic index: <25th centile (low risk), 25-50th centile (medium-low risk), 50-75th centile (medium-high risk), >75th centile (high risk).

An ability to generate personalized time-to-event curves that predict progression to CXL (Figure 2) could directly inform clinical decisions that benefit patient care. First, patients may better understand their own risk for progression and feel more confident in choosing their treatment options. Second, for both clinicians and patients, the prediction of progression may contribute to scheduling treatments, including prioritizing patients at high risk of early progression.

For example, patients at high risk with a 98% probability of progressing to CXL at 5 years could be offered CXL at the point of first diagnosis without waiting to demonstrate keratometric progression. Medium-risk patients may benefit from a period of clinician-led topographic monitoring. For the lowest-risk patients, optometry-led monitoring in the community may be sufficient. This risk stratification could be tailored to regions and reflect local needs and resources such as provision of monitoring services in regions with lower risk and greater capacity for CXL in areas with more high-risk patients.

Finally, when a decision is made to postpone CXL for further monitoring, the time-to-event curve can contribute to decisions on the scheduling of future follow-up reviews, with perhaps shorter time periods where the curve is steepest. Recommendations based on this model on clinical practice is yet to be evaluated.

Our study is subject to several limitations inherent to our data set. First, if patients had CXL at another hospital, this may not be reliably recorded in the source database. This could lead to a very small number of patients being included in the analysis who have already had CXL.

Second, ethnicity is a well-established risk factor for keratoconus and keratoconus progression,[27,30,31] but ethnicity is now an optional field at patient registration at our institution and this information was unavailable for approximately 50% of our data set. However, even when we restricted the data set to those with ethnicity records, it was not found to be a significant covariate.

Third, though the cohort used for univariable and multivariable analysis were identical, the number of eyes where all covariates were available was lower than for univariable analysis due to missing data. Finally, when we used multiple imputation to generate a multivariable model, ethnicity was still not found to be significant.

In the model fitting process, we chose to use a simple backward selection as opposed to the multivariate fractional polynomial method.[32] In our initial investigations, the results of multivariate fractional polynomial

yielded nonlinear functional forms of the covariates, and although this method may have slightly increased the predictive power of the prognostic model, the resulting hazard ratios would be very hard to interpret. In addition, we did not examine time-dependent effects for the covariates, which may provide a more accurate model fit, and future studies should examine this option. Finally, although no external validation data set was available, internal external cross validation allowed us to confirm that our model is generalizable across geographical regions.

In conclusion, we have fitted a prognostic model for progression of keratoconus to CXL that generates a time-to-event curve using age, Kmax, Front K1, and minimum pachymetry from time of presentation. Incorporation of a relatively small genetic data set does not improve the explained variation of our model. Personalized modeling of risk may improve patients' understanding of their condition and the need for CXL. Such a model may help better improve patients and aid clinician decision making to CXL to achieve better outcomes and judicious use of health care resources.

Funding/Support: Howard P. Maile is funded by a Moorfields Eye Charity PhD Studentship (GR001147). Nikolas Pontikos is funded by a Moorfields Eye Charity Career Development Award (R190031A). Moorfields Eye Charity is supported in part by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology. Stephen J. Tuft, Bruce D. Allan, and Daniel M. Gore acknowledge that a proportion of their financial support is from the Department of Health through the award made by the National Institute for Health Research to Moorfields Eye Hospital NHS Foundation Trust and University College London Institute of Ophthalmology for a Specialist Biomedical Research Centre for Ophthalmology.
Financial Disclosures: The sponsor or funding organization had no role in the design or conduct of this research. All authors attest that they meet the current ICMJE criteria for authorship. Howard P. Maile and J. Olivia Li contributed equally.

REFERENCES

1. Gordon MO, Steger-May K, Szczotka-Flynn L, et al. Baseline factors predictive of incident penetrating keratoplasty in keratoconus. *Am J Ophthalmol*. 2006;142(6):923–930.
2. Wittig-Silva C, Chan E, Islam FMA, Wu T, Whiting M, Snibson GR. A randomized, controlled trial of corneal collagen cross-linking in progressive keratoconus: three-year results. *Ophthalmology*. 2014;121(4):812–821.
3. Caporossi A, Mazzotta C, Baiocchi S, Caporossi T. Long-term results of riboflavin ultraviolet A corneal collagen cross-linking for keratoconus in Italy: the Siena eye cross study. *Am J Ophthalmol*. 2010;149(4):585–593.
4. O'Brart DPS, Chan E, Samaras K, Patel P, Shah SP. A randomised, prospective study to investigate the efficacy of riboflavin/ultraviolet A (370 nm) corneal collagen cross-linkage to halt the progression of keratoconus. *Br J Ophthalmol*. 2011;95(11):1519–1524.
5. Koller T, Mrochen M, Seiler T. Complication and failure rates after corneal crosslinking. *J Cataract Refract Surg*. 2009;35(8):1358–1362.
6. Gore DM, Leucci MT, Koay SY, et al. Accelerated pulsed high-fluence corneal cross-linking for progressive keratoconus. *Am J Ophthalmol*. 2021;221:9–16.
7. Salmon HA, Chalk D, Stein K, Frost NA. Cost effectiveness of collagen crosslinking for progressive keratoconus in the UK NHS. *Eye*. 2015;29(11):1504–1511.
8. Lindstrom RL, Berdahl JP, Donnenfeld ED, et al. Corneal cross-linking versus conventional management for keratoconus: a lifetime economic model. *J Med Econ*. 2020;24(1):410–420.
9. Godefrooij DA, Mangen MJJ, Chan E, et al. Cost-effectiveness analysis of corneal collagen crosslinking for progressive keratoconus. *Ophthalmology*. 2017;124(10):1485–1495.
10. Larkin DFP, Chowdhury K, Burr JM, et al. Effect of corneal cross-linking versus standard care on keratoconus progression in young patients: the KERALINK randomized controlled trial. *Ophthalmology*. 2021;128(11):1516–1526. doi:10.1016/j.ophtha.2021.04.019.
11. Vinciguerra R, Belin MW, Borgia A, et al. Evaluating keratoconus progression prior to crosslinking: maximum keratometry vs the ABCD grading system. *J Cataract Refract Surg*. 2021;47(1):33–39.
12. Shajari M, Steinwender G, Herrmann K, et al. Evaluation of keratoconus progression. *Br J Ophthalmol*. 2019;103(4):551–557.
13. Ozalp O, Atalay E. Belin ABCD Progression Display identifies keratoconus progression earlier than conventional metrics. *Am J Ophthalmol*. 2022;236:45–52.
14. Hashemi H, Panahi P, Asgari S, Emamian MH, Mehravaran S, Fotouhi A. Best Indicators for detecting keratoconus progression in children: a report from the Shahroud Schoolchildren Eye Cohort Study. *Cornea*. 2022;41(4):450–455.
15. Flynn TH, Sharma DP, Bunce C, Wilkins MR. Differential precision of corneal Pentacam HR measurements in early and advanced keratoconus. *Br J Ophthalmol*. 2016;100(9):1183–1187.
16. Flockerzi E, Häfner L, Xanthopoulou K, et al. Reliability analysis of successive Corneal Visualization Scheimpflug Technology measurements in different keratoconus stages. *Acta Ophthalmol*. 2022;100(1):e83–e90.
17. Kreps EO, Jimenez-Garcia M, Issarti I, Claerhout I, Koppen C, Rozema JJ. Repeatability of the Pentacam HR in various grades of keratoconus. *Am J Ophthalmol*. 2020;219:154–162.
18. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part IV: further concepts and methods in survival analysis. *Br J Cancer*. 2003;89(5):781–786.
19. Kato N, Negishi K, Sakai C, Tsubota K. Baseline factors predicting the need for corneal crosslinking in patients with keratoconus. *PLoS One*. 2020;15(4):e0231439.
20. Patrick Royston PL. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press; 2011.

21. Hardcastle AJ, Liskova P, Bykhovskaya Y, et al. A multi-ethnic genome-wide association study implicates collagen matrix integrity and cell differentiation pathways in keratoconus. *Commun Biol*. 2021;4(1):266.
22. Ding X, Guo X. *A survey of SNP data analysis*. Big Data Mining and Analytics. 2018;1(3):173–190.
23. Royston P, Parmar MKB. Flexible parametric proportional–hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002;21(15):2175–2197.
24. Baade PD, Royston P, Youl PH, Weinstock MA, Geller A, Aitken JF. Prognostic survival model for people diagnosed with invasive cutaneous melanoma. *BMC Cancer*. 2015;15:27.
25. Royston P, Parmar MKB, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer: construction and validation of prognostic model. *Stat Med*. 2004;23(6):907–926.
26. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med*. 2004;23(5):723–748.
27. Tuft SJ, Moodaley LC, Gregory WM, Davison CR, Buckley RJ. Prognostic factors for the progression of keratoconus. *Ophthalmology*. 1994;101(3):439–447.
28. Quartilho A, Gore DM, Bunce C, Tuft SJ. Royston—Parmar flexible parametric survival model to predict the probability of keratoconus progression to corneal transplantation. *Eye*. 2020;34(4):657–662.
29. Kato N, Masumoto H, Tanabe M, et al. Predicting keratoconus progression and need for corneal crosslinking using deep learning. *J Clin Med*. 2021;10(4):844. doi:10.3390/jcm10040844.
30. Pearson AR, Soneji B, Sarvananthan N. Does ethnic origin. *Eye*. 2000;14:625–628.
31. Georgiou T, Funnell CL, Cassels-Brown A, O'Conor R. Influence of ethnic origin on the incidence of keratoconus and associated atopic disease in Asians and White patients. *Eye*. 2004;18(4):379–383.
32. Royston P. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Continuous Variables*. John Wiley; 2008.